

Erasmus

Erasmus Student Journal of Philosophy

#6 | Juni 2014

Editorial

The sixth issue of the ESJP offers four excellent papers selected from a growing number of high-quality submissions. The journal is starting to benefit greatly from professors who urge their students to write not just to pass their courses, but also to be published in the ESJP. It is also benefiting from its growing popularity among students. Our aim being to further enrich the philosophical environment in which the students at our faculty develop their thinking, we are delighted to, for the first time, feature an essay that critically examines thoughts offered in a contribution from an earlier issue. The editorial board is proud to facilitate such discussions and hopes the papers published in our journal inspire other students to take part as well. We are grateful to all professors who have nominated papers and who have assisted the editorial board by taking the time to provide us with excellent comments and advice. This journal is only possible thanks to their continued effort.

The editorial board is currently going through a period of transition. The hard work of three guest editors who are Research Master students of the Erasmus Institute for Philosophy and Economics, Roel Visser, Elina Vessonen and Huub Brouwer, has proven to be invaluable for this issue. Their dedication and skill have deeply impressed me from the beginning of this issue's editorial process. This has been the last issue of Lydia Baan Hofman, who has been with us since the third edition. I would like to thank her for her support and for blessing the journal with her philosophical talent during her two years as an editor. Vivian Visser is also leaving the editorial board. Her willingness to serve as an excellent, impartial commentator on crucial papers and her hard work as a part of the design team has greatly contributed to the last two editions. Tim van Dijk, who has served as this issue's secretary, will be leaving as well. Both as an editor and as a secretary, he has clearly demonstrated himself to be irreplaceable. I greatly admire his keenness and command of language. As a secretary, he offered superb advice whenever needed and did much more work than could have been reasonably expected of him. Although I cannot foresee what will happen in the

future, for now, this has also been my last issue. I am incredibly grateful to Julien Kloeg, the previous editor-in-chief, for investing so much trust in me. He has been my patient mentor and dependable friend ever since I joined the editorial board. In his capacity as a member of the Advisory Board, he has helped out with the present issue by offering his advice whenever needed.

With us leaving, the time has come for others to lead the way. Huub Brouwer will be the editor-in-chief of the next issue. Already as a guest editor, he has shown remarkable ability, not only as an editor, but also as a writer. He is the author of one of the present issue's papers. I am positive that he will take the ESJP to greater heights and overcome the many challenges that lie ahead, such as the recruitment and training of new editors. In this, he will be assisted by Elina Vessonen, who will become the next secretary, and Roel Visser, who will work as an editor. Thijs Heijmeskamp, who has been a part of the editorial board from the very first issue, will continue his work as an editor. His experience and enthusiasm will be of great value to the new generation.

These opening lines are not complete without a word of thanks to the following people: the authors of the papers, for their willingness to revise and tolerate an endless amount of comments; dr. Bart Leeuwenburgh, dr. Patrick Delaere, prof. dr. Wiep van Bunge and prof. dr. Jack Vromen for serving as the jury of this year's Bayle Bokaal; Jasper van den Herik, for helping us out with the website and being there during crucial moments of the design process - we congratulate him with winning the Bayle Bokaal; Daan Gijsbertse, for his words of wisdom to the new editor-in-chief; Attilia Ruzzene, for contributing to the journal in various ways, and, last but not least, the readers of the ESJP, who will hopefully enjoy the fruits of our combined efforts.

Dennis Prooi
Editor-in-Chief

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

ISSN: 2212-9677

Editorial Board

Dennis Prooi (Editor-in-Chief)

Tim van Dijk (Secretary)

Lydia Baan Hofman

Thijs Heijmeskamp

Vivian Visser

Huub Brouwer (guest editor)

Elina Vessonon (guest editor)

Roel Visser (guest editor)

Advisory Board

Daan Gijsbertse

Myrthe van Nus

Julien Kloeg

Jasper van den Herik

Supervisory Board

prof. dr. J.J. Vromen

prof. dr. F.A. Muller

dr. P.J.J. Delaere



All work in this issue of the Erasmus Student Journal of Philosophy is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported License. For more information, visit <http://creativecommons.org/licenses/by-nc/3.0/>

Disclaimer

Although the editors of the Erasmus Student Journal of Philosophy have taken the utmost care in reviewing the papers in this issue, we cannot exclude the possibility that they contain inaccuracies or violate the proper use of academic referencing or copyright in general. The responsibility for these matters therefore remains with the authors of these papers and third parties that choose to make use of them entirely. In no event can the editorial board of the Erasmus Student Journal of Philosophy or the Faculty of Philosophy of the Erasmus University Rotterdam be held accountable for the contents of these papers.

In this issue

The sixth issue of the ESJP contains essays whose authors position themselves within ongoing academic debates in promising ways.

Do political theories have to conform with a plausible metaphysical theory? In *On Metaphysical Cases against Political Theories*, Manuel Buitenhuis answers this question with a resounding 'no'. In arguing for his case, he enters into a discussion with both Saul Smilansky and Willem van der Deijl.

In *Discussing Why Some Things Should Not Be for Sale*, Huub Brouwer provides us with an assessment of Debra Satz's moral limits to the market theory. This theory can be used to label markets we think of as morally problematic as 'noxious'. Brouwer argues that further elaboration of the framework on three issues would increase its usefulness for moral reflection on markets even further.

In *Would You Press a Button that Kills All Psychopaths?*, Nora Neuteboom modifies Causal Decision Theory in order to defend it against counterexamples formulated by Andy Egan. With the Two-Order Ratificationism Theorem thus obtained, she proves that there is a ratificationist account that delivers the right rational choice in the three-option cases pointed out by Anil Gupta.

Can Wittgenstein be seen as a part of the phenomenological movement? In *Wittgenstein vanuit continentaal fenomenologisch perspectief*, Myrthe van Nus demonstrates that this question can be answered positively when criteria formulated by Herbert Spiegelberg are applied. According to Van Nus, this positive answer implies that approaches considering Wittgenstein to be solely rooted in analytic philosophy fall short.

Table of contents

On Metaphysical Cases against Political Theories Manuel Buitenhuis	6-13	<u>5</u>
Discussing Why Some Things Should Not Be for Sale Huub Brouwer	14-21	
Would You Press a Button that Kills All Psychopaths? <i>The Two-Order Ratificationism Theorem of Causal Decision Theory</i> Nora Neuteboom	22-28	
Wittgenstein vanuit continentaal fenomenologisch perspectief Myrthe van Nus	29-39	

On Metaphysical Cases against Political Theories

Manuel Buitenhuis

'[...] the optimistic view – which seems to be at the heart of much metaphysical rationalism – [holds] that all good things must be compatible, and that therefore freedom, order, knowledge, happiness, a closed future [...] must be at least compatible, and perhaps even entail one another in a systematic fashion. But this proposition is not self-evidently true, if only on empirical grounds. Indeed, it is perhaps one of the least plausible beliefs ever entertained by profound and influential thinkers.'

Isaiah Berlin (2000: 117-118)

1. Introduction

It has become increasingly popular to criticize political theories from the perspective of the metaphysical debate on free will. Smilansky (1997) was one of the first to approach egalitarianism from this angle, and argued that the viability of different egalitarian theories is dependent on our solution to the free will problem, that is, the kinds of egalitarianism that are admissible depend strongly on the conception of free will they presuppose.

Following up on this endeavor with a more specific claim, Van der Deijl (2013) has used metaphysical arguments to criticize luck egalitarianism. Luck egalitarianism and its variations, a popular group of positions developed among others by Dworkin (2002), suggest that individuals should be compensated for negative effects due to those factors for which they have no responsibility. This compensation should be paid for at the expense of positive effects for which they were not responsible. Luck egalitarianism therefore requires a notion of responsibility that can separate factors for which we are responsible from factors for which we are not: it must separate luck from merit. Hard determinists cannot offer such an

account, because they see actions as necessarily following from physical states; compatibilists have tried, but according to Van der Deijl to no avail. An adequate notion of responsibility can be held by libertarians, but that position is unpopular because it requires the rejection of determinism. Van der Deijl concludes that, unless libertarianism can save the day after all, luck egalitarianism cannot distinguish luck from merit. In that case luck egalitarianism would collapse into outcome egalitarianism.

Supporting the arguments of Smilansky and Van der Deijl is the silent premise that metaphysical theory takes precedence over political theory, in the sense that a political theory should presuppose a plausible metaphysical theory. Kang has called this silent premise the 'Posteriority Thesis' (2003).

The Posteriority Thesis (PT) is intuitively appealing: it seems obvious that a political theory should not incorporate entities that cannot exist metaphysically. The consequences of accepting this thesis are, however, not nearly as intuitive since acceptance of PT forces us to dismiss notions that have not only proven to be useful in political philosophy, but that also seem to be morally significant. It also risks paralyzing political philosophy until certain metaphysical issues have been resolved. Although none of these points are in themselves sufficient reason to discard PT, they are reason enough to inquire into its foundation.

This paper subjects PT to a critical evaluation. Its main aim is to show that the truth of PT is often assumed to be trivial, while it has important consequences and while there are alternative approaches to political theory that can do without it. I will proceed by assessing whether it is correct to dismiss a political theory on the grounds of not being consistent with

plausible metaphysical theories about free will, as Smilansky and Van der Deijl have done. It should be added that it is explicitly not my intention to confirm or reject PT, nor to defend or reject any of the political theories discussed by Van der Deijl or Smilansky on non-metaphysical grounds. I merely wish to show that the influence of metaphysics in political theory is not necessarily as decisive as is sometimes thought.

The next section will introduce PT. The argument will then consist of three parts. Section 2 shows that PT is usually assumed rather than justified, and argues that PT has undesirable consequences. Section 3 shows that it is possible to do political philosophy without adopting PT, using the example of John Rawls. This example illustrates that metaphysical theories carry less, if any force for political theory if one takes moral beliefs as one's starting point.

2. Introducing the Posteriority Thesis

This section aims to clarify the notion of the Posteriority Thesis, and explains why it is important to subject it to scrutiny. The term 'Posteriority Thesis' was coined by Sung-Hak Kang (2003) in response to Smilansky (1997). Smilansky discussed egalitarianism as defended by Cohen. Cohen had argued that he could introduce his views on political theory without trespassing into the notoriously complicated free will debate. These are some of Smilansky's arguments:

'(i) Firstly, we cannot accept the assumption that people obviously and uncontentiously have the type of free will required by such an egalitarian conception as Cohen's. (ii) Secondly, the egalitarian case has much to gain from the clarification of the importance of the free will problem for questions of distributive justice (and more specifically, from the recognition of the implications of the lack of libertarian free will). However, one's judgment about the possible widespread attractiveness of egalitarianism would be very much influenced by one's position on the 'Compatibility Question' of the free will problem. (iii) Thirdly, the *form* of egalitarianism which will be considered adequate crucially depends on the position taken with regard to the free will problem.' (Smilansky, 1997: 159-160)

In his reaction, Kang uncovers the Posteriority Thesis as an implicit element of Smilansky's argument. While Kang admits that congruence between metaphysical and political theory is a theoretical virtue, he asserts that there is no 'metaphilosophical mandate' (2003: 108) that dictates (a) the necessity of congruence between the two, or (b) that metaphysics takes priority in achieving congruence.

Now let us try to define PT. It is important to notice that PT goes further than asserting the following:

1. If a political theory presupposes a metaphysical theory, it must be a plausible metaphysical theory.

PT claims more than (1) because Smilansky says that political theorists who *do not* supplement their egalitarianism with a metaphysical addendum on free will *should* do so: the 'adequate form of egalitarianism' after all *crucially depends* on one's position in the free will debate (see citation above). In other words, the possibility of putting metaphysical debates aside when doing political theory is rejected. Our working definition of PT should also incorporate that Smilansky assumes a hierarchy: our political theory should be adapted to fit our metaphysics, not the other way around. A better definition of PT would therefore be the following:

2. The only viable political theories are those that are grounded in a plausible metaphysical theory.

Statement (2) will serve as our definition of PT.

An important corollary of this definition is that a violation of PT is enough reason to dismiss a political theory. Van der Deijl exemplifies this in his argument against luck egalitarianism. To show this, let me first represent his argument.

According to Van der Deijl, luck egalitarianism requires 'substantive responsibility' in order to distinguish luck from merit. Substantive responsibility, a term borrowed from Scanlon (1999), refers to the kind of responsibility we carry for autonomous choices (that is, choices that are properly our own, rather than ascribed to us by others).¹ It can be contrasted with attributive responsibility, which regards choices that others attribute to us but are not autonomous.

Van der Deijl then asks which of the three positions in the metaphysical debate on free will entails substantive responsibility. Hard determinism, which holds that determinism is true and that, as a consequence, free will does not exist, does not entail substantive responsibility because hard determinism does not accommodate for individuals as autonomous agents: all human actions are the necessary consequence of the preceding physical states. The same problem holds for compatibilism. Compatibilists hold that some kind of free will can coexist with the truth of the determinist thesis. Since the problem for hard determinists was that determinism and substantive responsibility are not compatible, the same problem holds for compatibilists. Libertarianism, lastly, seems to do better with respect to substantive responsibility. Libertarianism rejects the determinist thesis and holds that free will exists, thereby allowing us to see some choices as autonomous. Libertarianism can, therefore, be compatible with substantive responsibility. The problem with libertarianism is, however, that it is considered an implausible position to begin with, since determinism carries too much appeal to be rejected.

Van der Deijl's conclusion is therefore that either libertarianism is true, or substantive responsibility cannot be saved. The latter possibility would imply that hard determinists and compatibilists cannot distinguish between luck and merit, and luck egalitarianism collapses into outcome egalitarianism.

I submit that Van der Deijl subscribes to PT defined as (2). To qualify, Van der Deijl must hold that a political theory is not viable if it is not supported by a plausible metaphysical theory. This seems to be the case when he says that luck egalitarianism must collapse into outcome egalitarianism if one's metaphysical theory does not offer us tools to separate luck from merit, as is the case for hard determinists and compatibilists. That compatibilism and hard determinism cannot metaphysically ground substantive responsibility should be a sufficient reason for dropping luck egalitarianism, unless one is a libertarian. By putting the argument this way, Van der Deijl rejects the option of justifying substantive responsibility on non-metaphysical grounds. This is evident in his response to Knight. Knight (2006) had suggested that one might not need a metaphysical basis to support one's political theory; moral intuitions could perhaps suffice.

Van der Deijl answers as follows:

'However, the implicit premise underlying [Knight's] argument is surely false: if our intuition goes against a metaphysical conclusion, the metaphysical conclusion is to be rejected. The point of drawing on arguments from the free will debate for the distributive justice discussion is exactly that it may show us that some of our intuitions may be wrong.'
(Van der Deijl, 2013: 28)

Because luck egalitarianism cannot be grounded in hard determinism or compatibilism, proponents of those metaphysical views cannot accept luck egalitarianism. If moral intuitions disagree with our metaphysical conclusions, the intuition should be rejected. Both show that Van der Deijl endorses PT: the only viable political theories are those that can be grounded in a plausible metaphysical theory.

3. Unwanted consequences of accepting the Posteriority Thesis

Recall that, according to Kang, there is no 'metaphilosophical mandate' demanding that metaphysics takes priority over political theory. This exposes a vulnerable link in the argumentation of Van der Deijl and Smilansky. A political theorist, let's call him McY might simply decide to reject PT, shrugging Van der Deijl's and Smilansky's works off. The argument might be something like this:

1. If PT has not been justified, it is arbitrary.
2. If PT is arbitrary, so are arguments that presuppose it.
3. If these arguments against (my preferred flavor of) egalitarianism are arbitrary, I can offset them by introducing an arbitrary criterion of my own (e.g. conformance to moral intuitions, justification by divine revelation)

Such an argument would, of course, not be very strong in itself, because it merely shifts the burden of proof. That PT is not justified does not necessarily mean it is arbitrary, since a justification might still be found

(contra 1); furthermore, the lack of explicit argumentation for PT does not come close to establishing the contrary (contra 3), regardless of whether that contrary is conceived as the independence or precedence of political theory. This argument only shows that PT is vulnerable; in order for McY's argument to carry any substantive weight, McY would have to argue that his criterion is more attractive than PT, or at least as attractive.

One way to do that might be to capitalize on the fact that PT may have undesirable consequences.

Firstly, acceptance of PT implies that some debates in political theory are impotent until metaphysical issues have been resolved (Fleurbaey, 2001). On the topic of luck egalitarianism, Knight mentions the demarcation of responsibility as an example of such a debate (2006: 814): accepting PT implies that luck egalitarianism can only be applied once consensus has been reached on a metaphysical notion of responsibility. Lacking such an answer, luck egalitarianism would not be able to progress. More generally, PT seems to suggest that we can only engage in political philosophy once we have settled all relevant metaphysical debates.

Secondly, PT has some counterintuitive consequences in the case of free will. Consider the following exchange: a defender of PT responds to McY's argument by holding that it would be inconsistent to use the concept of 'free will' in our political philosophy if our favored metaphysical theories suggest that free will does not exist.² McY then brings out the counterintuitiveness of PT with a variation: would it not be weird to do political philosophy without a notion of free will, since our moral intuitions about free will are so 'clear and distinct'? Why would I *a priori* reject a feeling that is so prominently present in every action I undertake? From this perspective, the problem is not that the justification of PT is weak. The problem is, instead, that such a weak justification is used to disqualify moral intuitions and beliefs that play an important part in our life. Even those who support the determinist thesis cannot deny that individuals have the feeling that they are actors that make decisions of their own. Much in the same vein, people who believe in hard determinism can nevertheless have strong moral intuitions about what is fair and what is not. Revising our political theories to comply with PT would classify a large class of appealing moral beliefs as irrelevant, because they cannot be grounded in our metaphysical theory.

Both consequences come together in a corollary of Van der Deijl's argument. If substantive responsibility cannot be saved by hard determinism or compatibilism, and the libertarian position is sufficiently implausible to not be considered, then adherence to PT requires us to reject all positions that presuppose substantive responsibility. This means that individuals cannot, under any circumstance, be seen as making autonomous choices. The consequences of that position are far-reaching. While this is not the place to discuss which positions and domains would be affected, the remainder of this section will try to give some examples.

Isaiah Berlin (2000) suggested that rejecting free will would force us to change the way we talk about human relations:

"Such expressions as 'I should not have done *x*', 'How could you have chosen *x*?' and so on, indeed the entire language of the criticism and assessment of one's own and others' conduct, would undergo a sharp transformation, and the expressions we needed both for descriptive and for practical-corrective, deterrent, hortatory purposes (what others would be open to a consistent determinist?) would necessarily be vastly different from the language which we now use." (Berlin, 2000: 108)

As a part of this change in vocabulary, accepting hard determinism would turn praise and blame to 'purely corrective or educational instruments, or confine them to aesthetic approval or disapproval' (Berlin, 2000: 102). For Berlin, praise and blame do not refer to the kind of praise and blame that can be attributed to someone under Scanlon's attributive responsibility, but only to praise and blame for autonomous actions: one is praised for the autonomous decision to save a drowning child, and one is blamed for the autonomous decision to steal. This kind of praise and blame cannot exist for hard determinists, since every action that is potentially praise- or blameworthy necessarily follows from preceding physical states. The only sense in which we can still speak of praise and blame is, according to Berlin, *aesthetic*. This implies that we can praise someone's bravery only in the same way that we praise someone's beautiful eyes. Both are, after all, the necessary consequence of physical causes; or conversely, neither follows from the autonomous choices of the individual.

What we have lost along the way is a *moral* perspective on praise and blame. As the citation foreshadowed, the loss of this kind of moral praise and blame seems to render statements and questions like ‘I should not have done x ’ or ‘How could you have chosen x ?’ empty. If I cannot make autonomous choices, my choice at moment t was ‘necessary’ in the same way that my regret at moment $t+1$ is; similarly, it seems to make little sense to blame you for choosing x if it was not your autonomous choice to begin with.

This point has a practical bearing on our justice system. Courts of law are often confronted with questions of responsibility. Is someone who committed a murder while sleepwalking responsible for his crime, even though he was not conscious (Lamme, 2011)? Is kleptomania an adequate excuse for theft, or should compulsion not count as a mitigating circumstance? If we reject autonomous choices on a metaphysical level and insist, through PT, that our political theory conforms to our rejection of the possibility of autonomous decision making, then our current way of treating these cases needs to be revised.

These examples offer insight into some of the consequences that the rejection of substantive responsibility may have. If we apply PT consistently, much more is lost than particular forms of egalitarianism. Of course, a defender of PT might decide to bite the bullet here. It just seems like quite a big bullet to bite, especially given the fact that PT is not independently justified.

4. John Rawls: the Plausibility of Political Theory without the Posteriority Thesis

This section shows that it is possible to do political philosophy without the Posteriority Thesis and that, consequently, metaphysical theory is not the only legitimate starting point of political philosophy. The lead is taken from Kang (2003). Arguing against Smilansky, Kang held that Rawls has shown how contradicting social values, such as equality and inequality, can be mediated by non-metaphysical criteria. While Rawls defends a limited amount of inequality, he does not do so on metaphysical grounds; that is, because a limited amount of inequality is entailed

by a metaphysical theory about the nature of human beings. Instead, he refers to a category of principles Kang calls ‘rational precepts’ (2003: 117). Examples of such rational precepts are economic analyses and instruments such as Pareto efficiency. Some limited level of inequality can be defended, for example, because it makes everyone better off economically. In deciding what level of inequality to allow in society, Rawls delegates decision-making to procedures on which society has, implicitly, agreed – not to metaphysical considerations.

Kang’s point is that there is more leeway between metaphysics and political theory than Smilansky is willing to admit. However, if we decide to take our cue from Rawls, there might be more lessons to be learned.

In *Justice as Fairness* (2001), Rawls emphasizes that his earlier works were often misinterpreted as built on the foundation of a particular set of philosophical assumptions about the nature of human beings. He argues that his views are to be interpreted as a political, rather than a philosophical theory.

‘Justice as fairness hopes to put aside long-standing religious and philosophical controversies and to avoid relying on any particular comprehensive view. It uses a different idea, that of public justification, and seeks to moderate divisive political conflicts and to specify the conditions of fair social cooperation between citizens.’ (Rawls, 2001: 29)

Rawls’ starting point differs from that of Smilansky and Van der Deijl. Smilansky and Van der Deijl inquire what political theory can be justified on the basis of a particular metaphysical theory; Rawls accepts a pluralist society, in which people have different norms, values and beliefs about the world, as his starting point. Rawls’ outset is to build a bridge between these different moral, religious and philosophical doctrines, rather than to convince anyone that a particular doctrine is wrong and should be exchanged for a different one. The ‘rational precepts’ that Kang refers to are one of the building blocks with which the bridge is built.

It is also interesting to note that Rawls explicitly states that the relevance of free will to political theory is to be drawn from its implicit use in the public sphere. If the metaphysical debate on free will has any importance for political theory, it is relevant because it influences the political views of individuals. Rawls thereby avoids having to deal with difficulties of the kind raised by Van der Deijl and Smilansky.

Richard Rorty's discussion of Rawls' views on the relation between politics and philosophy helps to bring out that Rawls' approach to political theory can coexist with hard determinism, compatibilism and libertarianism alike.

According to Rorty, Rawls' ability to cater to a pluralist society is made possible by the separation of metaphysics and political theory. For Rawls, questions regarding human nature are not relevant when we talk about the organization of our society. Rorty adds a historical twist to Rawls, suggesting that the beliefs and values that Rawls calls 'meanings embedded in political culture' can be seen in a historical and sociological perspective (Rorty, 2008: 181).

This emphasis on the historical contingency of the beliefs and values employed by Rawls should chime well with those who reject free will. Rorty points out, after all, that Rawls does not presuppose that these beliefs and values are the result of autonomous individual agency. He only requires them to be there. In the same vein, this emphasis on the historical contingency of beliefs and values should not prevent a Rawlsian framework from being accepted by compatibilists and libertarians.

The key point is that Rawls' view does not depend on a conception of individuals as either autonomous or not autonomous. This allows Rawls to be genuinely indifferent towards the metaphysical issue of free will: its outcomes do not matter to his method, because his starting point consists of moral dispositions *as experienced*, not *as caused*. This point is important, because it shows that Rawls' theory does not disagree with any specific position in the metaphysical debate – Rawls simply does not need them.

It is perhaps worthwhile to make a short detour to point out that I stray from Kang's own classification of Rawls when I say that he rejects PT. According to Kang, Rawls conforms to PT, but I think this is because

Kang applies his own criteria inconsistently. When discussing Rawls, Kang holds that Rawls *accepts* PT but *deviates* from it when there are rational precepts that justify doing so. Accordingly, Rawls *deviates* from the principle of income equality that follows from his hard determinist intuitions because rational precepts dictate that allowing limited inequality makes everyone better off. At the same time, PT was coined by Kang to represent the view that one's political theory should conform to one's metaphysical theory. In that case stating that Rawls *deviates* from PT when there are rational precepts that justify it seems to be contradictory. The rational precept that Kang uses to illustrate his point is, after all, an ethical statement: inequality is asserted to make 'everyone better off'. I contend Kang presents Rawls as deviating from the thesis that *metaphysics precedes ethics* on *ethical* grounds, which renders the thesis vacuous. PT would then hold that metaphysics precedes ethics unless there are ethical reasons to do otherwise.

That Kang's ascription of PT to Rawls renders the term vacuous can also be seen in a different way, since Kang offers no directions as to what can pass for a 'rational precept'. Kang admits that this question still needs to be answered, but might underestimate the consequence that it has for his categorization of Rawls. If the answer to the question is only a light constraint, then PT is more often ignored than upheld – and *vice versa* if the constraint is more severe. The only interpretation of PT that would be compatible with Rawls is a vacuous one, and consequently I interpret Rawls as rejecting PT.

It is now time to take all this back to the question of this paper. There are two points to be made here. Firstly, on a more general note, the example of Rawls shows that it is possible to justify a political theory on non-metaphysical grounds – in other words, that it is possible to engage in political theory if we do not uphold PT. In the case of Rawls, this is done by taking moral beliefs and dispositions as our starting point.

Secondly, relating specifically to Smilansky and Van der Deijl, the plausibility of a non-metaphysical justification of political theory calls into question whether they were right to dismiss a political theory that could not be justified metaphysically. It is, after all, possible to justify a political theory without any reference to metaphysical theories.

5. Conclusion

The aim of this paper has been to evaluate the Posteriority Thesis, as the claim that metaphysical theories take precedence over political theories so that only political theories that conform with a plausible metaphysical theory are admissible. PT is an implicit premise in Smilansky and Van der Deijl's arguments. In this paper I argued that it is questionable whether political theories can be decisively dismissed by such arguments.

The argument proceeded in two steps. Firstly, it was shown that PT itself is currently not supported by an independent argument, and that it has some significant undesirable consequences. Secondly, it was shown that political theories can be legitimately supported by non-metaphysical starting points. The example of Rawls illustrates that consistency with a particular metaphysical theory is not necessarily an attribute of a political theory, weakening the case for rejecting political theories on exclusively metaphysical grounds.

While the arguments of Smilansky and Van der Deijl hold when assuming that one's political theory must conform to one's metaphysical theory, this paper has tried to build credibility for the rejection of that assumption. The main aim of this paper would therefore be achieved when you have been convinced that the Posteriority Thesis is not a trivial assumption. Not only are its consequences counterintuitive and possibly paralyzing for political theory, but unquestioned acceptance of the Posteriority Thesis also risks closing us off from alternative approaches. There is more between heaven and earth than metaphysics.

Acknowledgements

I am grateful to Joris Vincent, Huub Brouwer, Tom Goodsir, Ingrid Robeyns and the anonymous reviewers of the ESJP for helpful commentary on a draft version of this paper.

Manuel Buitenhuis holds a Bachelor's degree from University College Utrecht, where he mainly studied philosophy and economics. He currently attends the Erasmus Institute for Philosophy & Economics (EIPE). He also follows a master programme in the History and Philosophy of Science at Utrecht University.

'On Metaphysical Cases against Political Theories' is an edited version of a paper that was originally written for the EIPE course 'Ethics & Economics' by prof. dr. Ingrid Robeyns.

Notes

1. There is a lot more to be said on this subject, but this definition should suffice for present purposes.
2. Van der Deijl exemplifies this when he holds that 'The point of drawing on arguments from the free will debate for the distributive justice discussion is exactly that it may show us that some of our intuitions may be wrong.' (Van der Deijl 2013: 28).

References

- Berlin, I. (2000) 'From Hope and Fear Set Free'. In: Berlin, I., *The Proper Study of Mankind: an Anthology of Essays*, 91-118. New York: Farrar, Straus and Giroux.
- Dworkin, R. (2002) *Sovereign Virtue: the Theory and Practice of Equality*. Cambridge: Harvard University Press.
- Fleurbaey, M. (2001) 'Egalitarian Opportunities'. In: *Law and Philosophy* 20(5), 499-530.
- Kang, S.-H. (2003) 'Free Will and Distributive Justice: A Reply to Smilansky'. In: *Philosophia* 31(1-2), 107-126.
- Knight, C. (2006) 'The Metaphysical Case for Luck Egalitarianism'. In: *Social Theory and Practice* 32(2), 173-189.
- Lamme, V. (2011) *De Vrije Wil Bestaat Niet*. Amsterdam: Uitgeverij Bert Bakker.
- Rawls, J. (2001) *Justice as Fairness*. Cambridge: The Belknap Press of Harvard University Press.

- Rorty, R. (2008) 'The Priority of Democracy over Philosophy'. In: Rorty, R., *Objectivity, Relativism, and Truth (Philosophical Papers: Volume 1)*, 175-196. New York: Cambridge University Press.
- Scanlon, T. (1999) *What We Owe to Each Other*. Cambridge: The Belknap Press of Harvard University Press.
- Smilansky, S. (1997) 'Egalitarian Justice and the Importance of the Free Will Problem'. In: *Philosophia* 25(1-4), 153-161.
- Van der Deijl, W. (2013) 'The Metaphysical Case against Luck Egalitarianism'. In: *Erasmus Student Journal of Philosophy* 4, 22-30.



Discussing Why Some Things Should Not Be for Sale

Huub Brouwer

'Your kidney for an iPad'

Advertisement slogan used by a Chinese kidney broker

Denis Campbell & Nicola Davison (2012)

Suppose your kidneys are failing. You might die before a donor kidney becomes available. If you are rich, there is a solution: buy a kidney on the black market in China, India or Pakistan. The kidney will cost you about \$200.000, of which the donor will receive only \$5.000. The remaining \$195.000 will go to the broker and the surgeon who made the transplant happen (Ibid.). Many of us think a competitive market in human kidneys would be troubling. This unease is reflected in the prohibition of kidney trade all around the world – except in Iran.

In *Why some things should not be for sale*, Debra Satz (2010) takes a critical look at markets that strike us as 'noxious' - markets that cause widespread discomfort or even revulsion (Ibid.: 3). She argues that most of these markets undermine our capacity to interact as equals. To guide discussions on the morality of markets, she develops a framework with four parameters that proxy for lacks in relational equality. These parameters are (1) weak agency, (2) extreme vulnerability, (3) extremely harmful outcomes to individuals, and (4) extremely harmful outcomes to society. A market that scores high on one or a few of these parameters can be deemed noxious.

Satz's framework backs our intuition that a competitive market in human kidneys is noxious (Ibid.: 204). This market would score high on all four parameters. People who sell a kidney are typically unaware of

the long-term health risks of doing so (weak agency) and they are likely to be destitute (extreme vulnerability). Kidneys could turn into debt collateral, making it harder for people who do not want to sell a kidney to get loans (extremely harmful outcome to individuals).¹ Finally, a kidney market could worsen existing class inequalities because only the rich would be able to buy kidneys and the poor would be supplying them (extremely harmful outcome to society).

Satz stresses that her framework is not intended as a blueprint that can be used to instantly recognize noxious markets. Users of her framework may have different opinions on the relative importance of the parameters and how high a market would need to score on them to label it 'noxious'. The project of this paper is to assess to what extent Satz's framework can be used as a guide for discussions on the morality of markets.

After starting off with a brief survey of Satz's moral limits to market theory, I will raise three issues. Firstly, since weak agency and extreme vulnerability occur often and do not always render a market noxious, it is not clear how these two parameters of Satz's framework could distinguish between markets that are problematic and markets that are not. Secondly, as the moral permissibility of markets depends to a large extent on how they are regulated, theoretical markets can only be assessed with a clearly specified regulatory framework in mind. Thirdly, it is very important to consider all available actual results of markets, as markets might turn out very differently in reality than they look on paper.

2. Satz's Moral Limits to Markets Theory

Satz (2010: 11) wrote her book for both theoretical and practical purposes. Her theoretical aim is to contribute a new perspective to discussions on equality by answering the questions *what type of equality is important?* and *how should this equality be achieved?* On a more practical level, she wishes to develop a framework that can be used to structure debates on the moral status of markets. This section briefly outlines Satz's theoretical contribution and her framework.

The theory

To see clearly what is novel about Satz's moral limits to markets theory, it is helpful to think about a competitive market in human kidneys again. According to Satz (Ibid.: 66), there are two important strands in the literature on equality that could help us to answer the question why this market is problematic: general egalitarianism and specific egalitarianism.

General egalitarians argue that markets may generate levels of social inequality that are unacceptable (Ibid.: 76). For them, a market in human kidneys would be problematic because it generates large inequalities: the rich will live longer because they are able to buy kidneys from the desperately poor. Banning the market would be inefficient, as competitive markets generate optimal economic results (Ibid.: 77). The solution is to reduce the inequality produced by the market through income redistribution.

A specific egalitarian on the other hand, claims that some goods should be distributed more equally than others (Ibid.: 79). Some specific egalitarians (for instance Michael Waltzer) even argue that certain goods should not be distributed by the market system at all, because selling them would degrade their meaning. Examples of such goods would be friendship, votes and Nobel Prizes. An argument that some specific egalitarians could make against a market in kidneys is that the human body and its constituent parts should not be commodified.

Satz thinks that general and specific egalitarians do not sufficiently consider the effects that markets have on the way people in a society

relate to each other.² To understand how markets and social structures are intertwined, she argues that we need to return to the works of classical political economists such as Adam Smith, David Ricardo and Karl Marx (Ibid.: 57).

According to them, markets can only function when they are socially embedded. To give one example, without the social institution of a legal system, the power of employers over employees would be potentially unlimited. If people could be enslaved or imprisoned by their employers, a competitive market in labor would be hampered (Ibid.: 43). Also, some markets shape us more than other markets (Ibid.: 40). Especially the labor market has a considerable impact on the types of persons we become.³ Sewing the same clothes in a sweatshop for fourteen hours a day, seven days a week could make a person passive and even servile – little more than an obedient appendix to the sewing machine.

The lesson we can learn from classical political economists is that markets shape the types of persons we are and the way we relate to each other. This is where Satz's concept of relational equality becomes important. She argues that because markets can shape the relations we have with each other, they can also undermine the conditions necessary for citizens in a society to 'interact as equals' (Ibid.: 95) and 'participate competently and meaningfully in democratic self-governance' (Ibid.: 101). 'Interacting as equals' here means that individuals do not have to beg or push others around to be able to make claims on one another. In most markets that we consider problematic, individuals do not interact as equals: there is a lack of equality of standing between parties before, during or after transactions take place (Ibid.: 93).

Satz takes a broader perspective than the general egalitarian: the problem with noxious markets is not only that they reinforce inequalities in income and wealth between the parties. Nor does she contest the claim made by specific egalitarians that certain goods cannot be sold because this would somehow degrade their meaning. The real problem for her is that some markets aggravate and perpetuate inequalities in background knowledge, level of education, social relations, and so on between their participants (Ibid.: 101).

The framework

To facilitate the discussion on whether a particular market is noxious or not, Satz develops a framework with four parameters that proxy for lacks in relational equality. Two of these parameters deal with the consequences of markets, the other two with their sources. A market is noxious when it scores high on one or more of the four parameters.

Satz takes extremely harmful outcomes to individuals to be outcomes that leave their basic interests unsatisfied. These basic interests refer to what is necessary to have a 'minimally decent' life (Ibid.: 95). An example of such a harmful outcome would be a stock market crash that completely evaporates someone's savings. Markets can also have extremely harmful outcomes to society. These occur whenever markets undermine the possibility for individuals to interact as equals (Ibid.: 95). Consider child labor. Children from poorer backgrounds will, on average, have to work more and will receive less education than children of richer parents. This way, the inequality in education level and income in a society is passed on from parents to the next generation.

Now, on to the source parameters. There is weak agency in a market whenever agents are not completely aware of the consequences of an action, or do not have full information about the good that is being exchanged (Ibid.: 96). An example would be a woman bearing someone else's child for money. She might feel very different about giving away the child after it has been born than at the moment when she signed the contract. Finally, there are extreme vulnerabilities in a market whenever there are large inequalities in bargaining power between market participants. This could be because one party has much fewer resources – be they wealth or capabilities – than the other; or because the supplier has market power. A case of extreme vulnerability would be a destitute woman who prostitutes herself to survive.

Using the Framework

Satz applies her framework in detail to five markets that people typically consider problematic: the market in surrogacy pregnancy, sex, child labor, voluntary slavery, and human kidneys. As Table 1 shows, these five markets set off all four of Satz's alarm bells for noxious markets. In that respect, these markets are extreme. A market scoring highly on just one parameter of the framework could already be problematic. Satz notes as a limitation of her approach that she does not define precisely what scores on parameters would be sufficient to label a market 'noxious'. This can lead to disagreements amongst users of her framework (Ibid.: 111).

Now, what to do if we agree that a market is noxious? Satz argues that a noxious market should not be banned automatically because the alternatives to banning the market might be worse. Forbidding child labor, for instance, could result in more child prostitution. Moreover, a change in regulation may already be sufficient to resolve the noxiousness of a market. A final reason for not banning problematic markets could be that banning is ineffective. Enforcing a ban on drugs, for example, is difficult. This could be a reason not to forbid a market in them.

Market	Extremely harmful outcomes	Sources
Surrogacy pregnancy	<i>Society:</i> perpetuates historical status inequality between men and women <i>Individual:</i> possibly threatens security of child, surrogacy mother could be left in destitution if contract is cancelled	<i>Weak agency:</i> possible future emotional costs to surrogate mother; unborn child's interests could be harmed <i>Vulnerability:</i> destitution/poverty
Prostitution	<i>Society:</i> perpetuates historical status inequality between men and women <i>Individual:</i> basic agency interests of individual harmed (especially in the case of streetwalkers)	<i>Weak agency:</i> pimps decide <i>Vulnerability:</i> destitution/poverty
Child Labor	<i>Society:</i> less productive workforce due to lack of education; higher costs of public health-care system; passive, ignorant citizenry <i>Individual:</i> undermining of future well-being and capacities of child	<i>Weak agency:</i> a child cannot make well-considered education decisions; unawareness of parents about benefits of education <i>Vulnerability:</i> destitution/poverty

Market	Extremely harmful outcomes	Sources
Voluntary slavery	<i>Society:</i> perpetuation of social inequalities that hamper capitalism <i>Individual:</i> extreme dependency on employer, inhibition to development of individual	<i>Weak agency:</i> illiteracy and innumeracy <i>Vulnerability:</i> destitution/poverty (no protection from seasonal fluctuations in income)
Kidney markets	<i>Society:</i> worsening of existing class inequalities <i>Individual:</i> potentially being forced to sell kidneys in case of destitution (kidneys could become collateral)	<i>Weak agency:</i> unawareness of potential health consequences <i>Vulnerability:</i> destitution/poverty

Table 1: Satz's assessment of noxious markets in *Why Some Things Should Not Be For Sale*

3. The discriminatory power of the source parameters

There are noxious markets that score high on only one or a few parameters of Satz's framework. An interesting case is when a market just scores high on the source parameters, but has no extremely harmful outcomes. Weak agency and extreme vulnerability occur often and they do not always make a market morally objectionable – even when they occur in high degrees (see Satz, 2010: 97; Claassen, 2012; Kachra, 2011). An important question is therefore: when is a market with only weak agency and/or extreme vulnerability noxious?

It is easily established that weak agency does not always make a market noxious. Consider the market in fast food. At least some consumers in this market will suffer from weak agency: they are not fully aware of the long-term health risks of eating fast food or do not know about the range of other food options available. Still, not many people think of the fast food market as morally objectionable. The same goes for markets with extreme vulnerabilities. Few people would deem a market in fruits with honest prices but some very poor buyers problematic.

In a review of Satz's book, Claassen (2012: 590) claims that weak agency and extreme vulnerability only render a market noxious when they lead to extremely harmful outcomes. This would mean that the outcome parameters in Satz's framework do all the discriminatory work. The source parameters would just be common causes of extremely harmful outcomes. We could leave them out of the framework and still label exactly the same markets 'noxious'.

This seems quite plausible at first glance. In all markets that Satz discusses in detail (see Table 1), weak agency and extreme vulnerability cause harmful outcomes. However, Satz does mention examples of cases where weak agency and extreme vulnerability make a market noxious even when there are no extremely harmful outcomes. For weak agency, she refers to markets that 'target young children', and markets 'whose products are based on deception' (2010: 97). For extreme vulnerability, she notes that markets in which people sell their assets at very low prices out of desperation are wrong, even if the sale does not result in extremely harmful outcomes (Ibid.).

Thus, Satz does mention that it is possible that markets are morally objectionable if there only is weak agency and/or extreme vulnerability. However, her framework does not seem to be developed to an extent that one can recognize such markets with it. Given that weak agency and extreme vulnerability occur often, it would be very helpful for discussions on the moral status of markets to have a better understanding of why they sometimes make markets noxious.

4. Regulatory context matters

Regulation and the noxiousness of markets are closely intertwined. Frequently regulation can resolve the problems in a market (Ibid.: 104). It even seems to be the case that in absence of regulation, almost any market can be noxious (Fleurbaey, 2011).⁴

An example could help to see this more clearly: suppose cars were not invented yet.⁵ Then, someone - call her Pauline Daimler - invents them. Many people will be shocked by the thought of cars: collisions can result in death (extremely harmful outcomes to individuals), drivers are largely dependent for their safety on other drivers whose capacities for driving they cannot establish (weak agency), and poorer car owners cannot afford the safer, heavier cars of the rich (vulnerability).

Without regulation, it seems that Satz would have to label a market in cars 'noxious'. With regulation, this is no longer the case. If there are clear traffic rules enforced by the police, a minimal safety standard for cars, and obligatory courses to ensure that drivers know the traffic rules and are capable of maneuvering their vehicles safely, the car market would not score so high on Satz's framework anymore. Regulation *can* make the difference between a market being noxious or not. This does not mean that regulation is a panacea for the noxiousness of markets, however. Some markets are problematic by their very nature. The already mentioned advertisement targeted at children would be an example.

Now, what to do with the knowledge that regulation and the moral status of markets are closely intertwined? When Satz applies her framework to markets in the third part of her book, she always asks two questions. First: is this market morally objectionable? And only when a market is found to be noxious: can regulation resolve its noxiousness? This strategy works well for markets that exist in reality, because then a regulatory context is already present.

However, many of the markets Satz considers are merely theoretical. There are, for instance, no nationwide markets in surrogacy motherhood or human kidneys (except for Iran in the latter case). Asking whether these markets are noxious before thinking about a regulatory framework for them seems to be unproductive. After all, even the market in cars and food could be noxious in absence of regulation.

5. Actual outcomes are surer than expected ones

A related worry is that it is very difficult to know what a theoretical market will be like when actualized. If one is unsure about how a market will turn out in reality, how can the market's moral status be assessed?

This question ties in with a classic problem in ethics. There is a debate in consequentialism on how moral knowledge – knowledge of whether an action is right or wrong – can be obtained (Shafer-Landau, 2010: 120). The central question is: how should we assess the moral status of an action when not all of its results have occurred yet?

Two answers are relevant here.⁶ Either you cannot tell, because the rightness of an action depends on its actual results. Or you can guess, because the rightness of an action can be inferred from its expected results. The main drawback of the first answer is that you do not have moral knowledge in cases where not all the results of an action have occurred yet. The main drawback of the second answer is that your expectations could turn out to be wrong.

A simple example can be of help here. Suppose you think that, in theory, there might be a risk that surrogacy motherhood would reinforce negative gender stereotypes. If you think that you cannot acquire moral knowledge without actual results, the only way to decide whether a market in surrogacy pregnancy would indeed reinforce negative gender stereotypes is to try this market. If you can base your judgment on expected outcomes, on the other hand, this would not be required.

In a review of Satz's (2010) book, Steiner (2013) argues that there is too large a possibility of making wrong judgments about markets that only exist in theory. According to him, one simply has to accept that it is impossible to assess the moral status of theoretical markets. I think such a strict stance would be a loss. It is as likely for theoretical markets to require moral assessments as it is for actual markets. Consider the privatization of public transport, electricity supply, and postal services during recent decades in many Western European countries. Moreover, if a market always has to be tried before it could be forbidden, much harm might occur that could have been anticipated.

It is good to be aware, however, that there always is a risk of having wrong expectations about how a market will turn out in reality. This means that the user of Satz's framework needs to consider any available empirical evidence on the markets assessed. Sometimes there is a lack of such attention to empirical studies in Satz's book.

For starters, Satz argues that prostitution is noxious *inter alia* because it reinforces negative gender stereotypes, according to which women are 'objects for male use' (Ibid.: 149). She does not, however, try to establish whether this is indeed what happens in countries such as Argentina, Germany, the Netherlands, Mexico and the United Kingdom, where prostitution is legal in varying degrees (Global News and The Canadian Press, 2013). Similarly, in her discussion of the potential noxiousness of a market in human kidneys, Satz does not consider the various empirical studies done on the kidney market in Iran – she just mentions in a footnote that Iran is the only country in the world with a legal market in kidneys (Ibid.: 233). Such empirical studies would give her some actual results with which she can contrast her expectations.

6. Conclusion

The aim of this paper was to assess to what extent Satz's framework can be used as a guideline for discussions on the moral status of markets. As Satz demonstrates in her book, the framework is successful in the sense that many markets we think of as problematic also score highly on one or more of its parameters. Nevertheless, I have argued in this paper that some further elaboration of the framework and how to use it would be helpful on three issues.

First of all, some markets are noxious just because they score high on the source parameters of Satz's framework. Such cases are interesting, because weak agency and extreme vulnerability occur often and do not always render a market noxious. Consequently, it would be desirable to have a criterion that tells us when weak agency and extreme vulnerability are problematic and when they are not.

Also, almost any market could be noxious without regulation. When considering theoretical markets, it is therefore unproductive to first ask whether the market is noxious and only afterwards whether this noxiousness can be resolved through regulation. Even the commonplace markets in food and cars would be problematic without regulation.

Lastly, when assessing the moral status of a market, it is important to consider its actual outcomes as much as possible. There are many ways in which a market might turn out differently in reality from how it looks on paper. To avoid making erroneous assessments of markets, it is therefore important to consider all available empirical evidence.

Any guide for discussion is incomplete by definition. This also goes for Satz's guide for moral reflection on markets. The purpose of this paper was to point at some details where further elaboration of her framework would facilitate discussions on the moral status of markets even more. The devil is always in the details; it is good to know where it is hiding.

Acknowledgements

I am grateful to Professor Ingrid Robeyns and Roel Visser for helpful comments and suggestions. The quality of this paper improved significantly as a result of the suggestions of anonymous Erasmus Student Journal of Philosophy referees.

Huub Brouwer (1989) completed an MSc in Business and Economics at the Stockholm School of Economics, studied political theory during an exchange at Sciences Po Paris, and obtained a BA in Liberal Arts and Sciences at University College Utrecht. He is currently enrolled in the MA program in Philosophy and Economics at Erasmus University Rotterdam. His main research interests are in the areas of Ethics and Economics and the Epistemology of Economic Modeling.

'Discussing Why Some Things Should Not Be for Sale' was written for the course 'Ethics and Economics' taught by Prof. Dr. Ingrid Robeyns

Editorial note:

Since this essay was written by a member of the editorial board of the Erasmus Student Journal of Philosophy, it was subject to a more extensive review procedure. For more information, see <http://www.eur.nl/fw/english/esjp/submissions>.

Notes

1. Satz (2010: 200) refers to a study by anthropologist Lawrence Cohen, in which he notes that kidneys – especially those of women – are already used as debt collateral in certain regions of India.
2. For a summary of Satz's critique of general egalitarians and specific egalitarians, see Claassen's (2012) review of her book.
3. Adam Smith already pointed this out in his *Inquiry into the Nature and Causes of the Wealth of Nations* (1776/1904). He writes that factory work can make workers as 'stupid and ignorant as it is possible for a human creature to become' (Ibid.: V.1.178).
4. Satz (2010: 99) acknowledges this, but then does not apply it in her discussion of specific markets in the third part of her book.
5. Fleurbaey (2011) mentions the food market as an example of a market that would score high on multiple parameters in Satz' framework absent regulation.
6. I ignore 'proximate consequentialism', which uses the legal notion of proximate causes to determine the moral status of actions (Sinnott-Armstrong, 2012). It seems relevant to neither Satz' nor Steiner's position.

References

- Anderson, E. (2012) 'Why some things should not be for sale: The moral limits of markets Debra Satz'. In: *New Political Economy* 17(2), 239-242.
- Arneson, R. (2013) 'Egalitarianism', *The Stanford Encyclopedia of Philosophy* (Summer 2013 Edition), Edward N. Zalta (ed.), Retrieved from: <http://plato.stanford.edu/archives/sum2013/entries/egalitarianism/> [05 January 2014]
- Claassen, R. (2012) 'Why Some Things Should Not Be For Sale: The Moral Limits of Markets, by Debra Satz'. In: *Business Ethics Quarterly* 22(3), 589-601.

- Campbell, D. & Davison, N. (2012) "Illegal kidney trade booms as new organ is 'sold every hour'". Retrieved from: <http://www.theguardian.com/world/2012/may/27/kidney-trade-illegal-operations-who> [5 June 2014]
- Fabre, C. (2011) 'Debra Satz, Why some things should not be for sale'. In: *Ethics* 121(2), 469-475.
- Fletcher, C. (2011) 'Why some things should not be for sale: The moral limits of markets'. In: *Journal of Markets and Morality* 14(1), 211-213.
- Fleurbaey, M. (2011) 'Why some things should not be for sale, Oxford University Press, 2010'. In: *Oeconomia* 1(3).
- Global News and The Canadian Press (2013) 'Prostitution laws around the world'. Retrieved from: <http://globalnews.ca/news/128029/at-a-glance-prostitution-laws/> [2 February 2014]
- Heath, J. (2011) 'Review of Debra Satz's Why some things should not be for sale. Oxford: Oxford University Press, 2010, 252 pp'. In: *Erasmus Journal for Philosophy and Economics* 4(1), 99-107.
- Kachra, K. (2011) 'Review - Why some things should not be for sale'. In: *Metapsychology* 15(3).
- Lehman, C. (2012) 'The higher bealism: The moral limits of markets'. Retrieved from: <http://www.thenation.com/article/168240/higher-bealism-moral-limits-markets> [5 January 2014]
- Maskivker, J. (2011) 'Why some things should not be for sale: The moral limits of markets'. In: *Perspectives in Politics* 9(4), 933-934.
- Satz, D. (2010) *Why some things should not be for sale: The moral limits of markets*. New York: Oxford University Press.
- Schmidtz, D. (2011) 'Why some things should not be for sale: The moral limits of markets'. In: *Journal of Philosophy* 108(4), 219-223.
- Shafer-Landau, R. (2010) *The fundamentals of ethics*. New York: Oxford University Press.
- Sinnott-Armstrong, W. (2012) 'Consequentialism', *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), Edward N. Zalta (ed.), Retrieved from: <http://plato.stanford.edu/entries/consequentialism/> [5 January 2014]
- Smith, A. (1904) *An inquiry into the Nature and the Causes of the Wealth of Nations* (5th Edition). London: Methuen & Co.
- Steiner, P. (2011) 'Contesting markets'. In: *European Journal of Sociology* 52(3), 546-548.



Would You Press a Button that Kills All Psychopaths?

The Two-Order Ratificationism Theorem of Causal Decision Theory

Nora Neuteboom

1. Introduction

Paul is debating whether to press a button that 'kills all psychopaths'. Surely it would be much better to live in a world without psychopaths, Paul reasons. However, at the same time Paul is quite confident that only a psychopath would press such a button. Paul prefers living in a world with psychopaths to dying. What should Paul do?

In this situation it seems irrational to press the button and rational to refrain from doing so: if Paul would press he would be a psychopath and hence die.

However, in a recent paper Egan (2007) points out that the most sophisticated current version of decision theory, commonly called Causal Decision Theory, actually gives the irrational recommendation to press the psychopath button in this example.

How can this be? Decision theory may be pursued with the aim of finding out how decisions ought to be made in order to maximize one's utility.¹ In the standard view, rational choice is defined as the process of determining what options are available and then choosing the most preferred one. Within the rational actor model, which is the standard for Rational Choice Theory (RCT), the substance of the rational action concept is the maximization of individual utility. The individual can choose between alternative courses of action, taking into account individual preferences and states of nature. The rationality of a decision is evaluated by taking into account in what measure the results of the decision have succeeded in maximizing utility (that is, satisfying individual preferences) under some specific contextual constraints (states of nature).

There is an ongoing debate between two camps of decision theorists, namely, the proponents of Evidential Decision Theory (EDT) and the proponents of Causal Decision Theory (CDT). The difference between the two decision theories consists in how they compute the relative value of actions. Roughly speaking, EDT says to do the thing you would be happiest to learn that you did, and CDT says to do the thing that is most likely to bring about the best results (Egan, 2007: 93). In this paper, I try to contribute to the debate between causal and evidential decision theorists by solving the latest counterexamples that burden CDT by proposing a revisionist account.

In part one I begin by briefly introducing Egan's 2007 paper which shows that there are fatal problems to using both EDT and CDT for choosing rational options. Thereafter I spell out the Ratificationism Theorem (RT). Ratificationism requires the chosen act, A, to have an estimated desirability at least as high as any of the alternative choices on the hypothesis that one's final decision will be to perform A (Jeffrey, 1983: 19). I reconfirm that RT is not sufficient for solving all disputes between CDT and EDT. A problem with ratificationism is that in some cases there are no ratifiable options, but some options still seem rational. To solve this problem, Egan considers a lexical version of Ratificationism. The Lexical Ratificationism Theorem (LRT) recommends at least one option, even in cases where no option is ratifiable. I demonstrate, following Gupta, that also LRT can be refuted with a counterexample. In part two of the paper I proceed to develop an alternative ratificationist account, the Two-Order Ratificationism Theorem (TORT). I show that TORT does give us the correct results to two recently posed (and still unresolved) counterexamples to LRT.

PART I

1.1 Causal and Evidential Decision Theory

In recent philosophical debates the most prominent rival to CDT is EDT. EDT upholds that the best action is the one which, *conditional* on you having chosen it, gives you the best expected outcome. CDT maintains that the expected utility of actions should be *unconditionally* evaluated with respect to their potential consequences. CDT enjoins us to do whatever has the best expected outcome, holding fixed our initial views about the likely causal structure of the world (Egan, 2007: 94-96).

In his 2007 paper *Some Counterexamples to Causal Decision Theory*, Egan shows that there are fatal problems to both EDT and CDT. EDT argues for a policy of performing the action with the greatest evidential value, rather than the action with the best-expected causal upshot (2007: 93-96). CDT requires that the expected utility of an action is unconditionally evaluated with respect to its potential consequences (2007: 96-102). However, both EDT and CDT are mistaken in some cases according to Egan. To solve this, he modifies the CDT account by proposing RT.²

1.2 Ratificationism Theorem

(RT)

It is rational to perform an action A, iff,

1. *A is ratifiable, and*
2. *There is no other ratifiable option with greater VAL_{CDT} than A.*

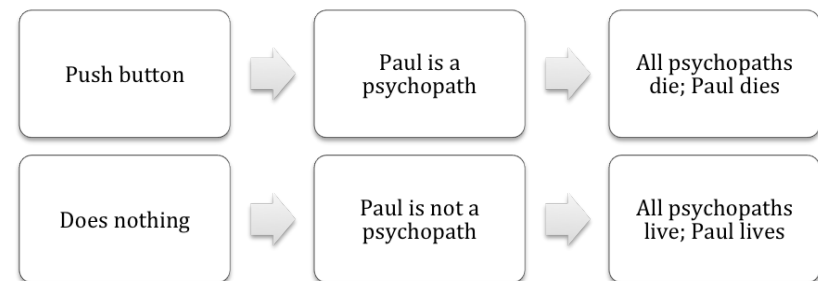
Option A is ratifiable if, and only if, there is no alternative B such that the value of B ($VAL_{CDT}(B)$) exceeds the value of A ($VAL_{CDT}(A)$) on the supposition that A is decided upon. Therefore, it is rational to decide upon an option A if, and only if, A is the only ratifiable option.

Egan (2007: 107) promotes RT as a kind of adjusting principle; whenever basic CDT would get it wrong, RT should help to get to the right answer.

But it seems that RT is not useful for all cases. Let us go back to where we started our journey with, the example of the psychopath:

Paul is debating whether to press the ‘kill all psychopaths’ button. It would, he thinks, be much better to live in a world with no psychopaths. Unfortunately, Paul is quite confident that only a psychopath would press such a button. Paul prefers living in a world with psychopaths to dying.

I represent Paul’s choices below:



If Paul would press the button, he would be a psychopath and hence die. Pressing is the irrational thing to do since Paul prefers living to dying. In order to have an account of decision theory that gives us the utility-maximizing guidance on rational acts, we want our decision model to advise us to not press the psychopath button.

CDT advises us to unconditionally evaluate the outcome. In this example, the fact that Paul is a psychopath (if he presses the button) is an unfortunate condition of pressing the button. But CDT does not take this condition into account and therefore encourages pressing the button. Also RT cannot help CDT from falling afoul, since neither refraining from pressing the button nor pressing the button is ratifiable. When Paul becomes convinced that he will choose to refrain, he will become quite confident that he is not a psychopath, and pressing will look better than refraining. Thus the option ‘not pressing’ is not ratifiable because there

is an alternative option, ‘pressing’, such that the value of that alternative option ($VAL(PRESSING)$) exceeds the value of the initial option ($VAL(NOT-PRESSING)$), on the supposition that the initial option ‘not pressing’ is decided upon. The same holds for the option ‘pressing’. The moment that Paul becomes convinced that ‘pressing’ is the best option, the value of the option ‘not pressing’ ($VAL(NOT-PRESSING)$) exceeds the value of ‘pressing’ ($VAL(PRESSING)$) since as soon as Paul becomes aware of the fact that he prefers to press, he might be a psychopath himself. So when Paul chooses to press, the value of not pressing will exceed the value of pressing since he has a good reason to think that he is actually a psychopath.³ It seems that no option is ratifiable, and thus RT does not give us any helpful guideline in this example.⁴

1.3 Egan’s Lexical Ratificationism Theorem

Since Paul is quite sensitive about dying, we have the obligation to fix this problem. To do so, Egan (2007: 111) considers an adaption of RT, LRT.

(LRT)

It is rational to decide upon an option A iff,

1. *A is ratifiable, and there is no other ratifiable option with higher VAL_{EDT} than A,⁵ or*
2. *There are no ratifiable options, and no other (unratifiable) option has higher VAL_{EDT} than A.*

Option A is ratifiable if, and only if, there is no alternative B such that value of B ($VAL_{CDT}(B)$) exceeds the value of A ($VAL_{CDT}(A)$) on the supposition that A is decided upon. Therefore, it is rational to decide upon an option A if, and only if, A is the only ratifiable option.

Step one advocates to order by ratifiability – that is, if A is ratifiable and B is unrati- fiable, then A is to be preferred over B. Step two advocates that within each of the two groups, we should order by VAL_{EDT} .⁶ Thus, it seems that LRT advances CDT by taking some elements from EDT for rational decision-making.

LRT recommends at least one option even in cases where no option is ratifiable. For example, it yields the rational recommendation not to press the psychopath button. Step one is indecisive since we already saw that neither refraining from pressing the button nor pressing the button is ratifiable. In step two we refer to the EDT principle. EDT says that the rational action is the one such that your expected utility, conditional on you performing it, is greater than the expectations conditional on you performing any other action. So conditional on the fact that *if* you press you would be a psychopath, EDT prescribes not to press. And that seems to be the rational thing to do.

1.4 A refutation of the Lexical Ratificationism Theorem

LRT seemed the way to go until Gupta (Egan, 2007: 112) came up with a decisive counterexample of the Three-Option Smoking Lesion:⁷

Paul is deciding whether to smoke. Paul has three options: Smoke cigars, smoke cigarettes, or refrain from smoking altogether. Call these options CIGAR, CIGARETTE, and NO SMOKE. Due to the ways that various lesions tend to be distributed, it turns out that cigar smokers tend to be worse off than they would be if they were smoking cigarettes, but better off than they would be if they refrained from smoking altogether. Similarly, cigarette smokers tend to be worse off than they would be when smoking cigars, but better off than they would be when refraining from smoking altogether. Finally, non-smokers tend to be best off refraining from smoking.

Assume:

ψ_1 = smoking cigars

ψ_2 = smoking cigarettes

$\neg\psi$ = not smoking

I represent the example formally in Table 1 below.

ψ : Smoker	$\neg\psi$: Non-smoker
$VAL(\psi_1) > VAL(\neg\psi)$	$VAL(\neg\psi) = VAL(\psi_1)$
$VAL(\psi_2) > VAL(\neg\psi)$	$VAL(\neg\psi) = VAL(\psi_2)$
$VAL(\psi_1 \psi_2) > VAL(\psi_2 \psi_1)$	
$VAL(\psi_2 \psi_1) > VAL(\psi_1 \psi_2)$	

Table 1: The values of the Three-Option Smoking Lesion

Following LRT, step one tells Paul that there is one ratifiable option, not smoking ($\neg\psi$), and there are two unratable options (ψ_1, ψ_2). Both ψ_1 and ψ_2 are unratable since smoking cigarettes is very good evidence that you would be better off smoking cigars and vice versa. Not smoking ($\neg\psi$), however, is ratifiable because it is good evidence that you would be best off not smoking. Thus, based on step one and two in LRT, Paul should not smoke. But this seems irrational: if Paul wants to smoke cigars or cigarettes, one thing Paul knows for sure is that not smoking is not the way to go.⁸

Egan thinks that this example is fatal for LRT:

‘No ratificationist account will be able to deliver the right results in the sorts of three-option cases that Gupta has pointed out. The real importance of the Gupta cases is not that they refute lexical ratificationism— it’s that they refute every form of ratificationism’ (Egan, 2007: 113).

In Part II, I will prove that this is not true by providing an alternative account of LRT that does give the rational solution to Gupta’s counterexample.

PART II

2.1 An adapted version of the Lexical Ratificationism Theorem

So let us look at what element in the Three-Option Smoking Lesion really causes the structural failure of LRT.

Paul has three options. Option ψ_1 is unratable because, conditional on choosing option ψ_1 , option ψ_2 looks better than option ψ_1 – and the same applies to option ψ_2 . Option $\neg\psi$ is ratifiable because, conditional on choosing option $\neg\psi$, option $\neg\psi$ looks better than either ψ_1 or ψ_2 . However, conditional on choosing either one of the options ψ_1 or ψ_2 , option $\neg\psi$ looks very bad. What seems clearly irrational, for Paul, who finds himself deciding on either ψ_1 or ψ_2 , is to perform action $\neg\psi$ on grounds of its ratifiability. What we need is an account that considers the fact that if Paul finds himself in a situation in which he wants to choose between ψ_1 and ψ_2 , the ratifiability of option $\neg\psi$ does not hold as binding anymore. So considering Paul’s desire to smoke, option ψ_1 and ψ_2 should also be ratifiable.⁹ From this line of thought it seems that we need to consider two sorts of ratifiability. To incorporate two sorts of ratifiability in the account, I develop the Two-Order Ratificationism Theorem version of CDT:

(TORT)

It is rational to decide upon an option A iff,

1. *A is first-order or second-order ratifiable and*
2. *There are no other first-order or second-order ratifiable options with higher VAL_{EDT} than A, or*
3. *There are no ratifiable options, and no other (unratifiable) option has higher VAL_{EDT} than A.*

An option A is first-order ratifiable if, and only if, there is no alternative B such that $VAL_{CDT}(B)$ exceeds $VAL_{CDT}(A)$ on the supposition that A is decided upon.

An option A is second-order ratifiable if, and only if, there is no other first-order ratifiable alternative B such that $VAL_{CDT}(B)$ (on the supposition that B is decided upon) exceeds $VAL_{CDT}(A)$ (on the supposition that A is decided upon).¹⁰

To demonstrate the above TORT principle, I will examine the Three-Option Smoking Lesion again. The option not to smoke ($\neg\psi$) is obviously a first-order ratifiable option; there is no alternative (ψ_1 or ψ_2) such that the $VAL_{CDT}(\psi_1)$ or $VAL_{CDT}(\psi_2)$ exceeds the $VAL_{CDT}(\neg\psi)$ on the supposition that Paul is a non-smoker ($\neg\psi$). Then we proceed by arguing that both option ψ_1 and option ψ_2 are second-order ratifiable. Option ψ_1 is a second-order ratifiable option since there is no other first-order ratifiable alternative ($\neg\psi$) such that

the $VAL_{CDT}(\neg\psi)$ exceeds $VAL_{CDT}(\psi_1)$ on the supposition that Paul chooses to smoke. Since Paul already decided to smoke, the value of ψ_1 exceeds the value of $\neg\psi$ in the example. The same applies to option ψ_2 . So all options are either first-order or second-order ratifiable. Therefore Paul has to choose the ratifiable option that has the highest VAL_{EDT} . If ψ_1 or ψ_2 is ruled out in favor of $\neg\psi$, it is due to a higher VAL_{EDT} of $\neg\psi$. Hence, unlike LRT, TORT does not rule out ψ_1 or ψ_2 in the Three-Option Smoking Lesion on grounds of $\neg\psi$'s ratifiability, since all options are ratifiable. TORT recommends a rational choice in the Three-Option Smoking Lesion.

One could argue, however, that TORT is not really a pure ratifiability account. Since I introduce a second-order ratifiability, the content of the notion ratifiability changes, thereby changing its relation with rationality. However, one should notice that ratifiability does not necessarily imply rationality. In other words, not all ratifiable options are rational, even though all non-ratifiable options are irrational. What a ratifiability account should do is to rule out the irrational options for the agent. As Egan (2007: 108) points out: 'being unrati-fiable is sufficient for being ruled out as a rational option'. In the above case, Paul choosing not to smoke ($\neg\psi$) is not ruled out as an irrational option immediately. However, we cannot say it is rational either. What matters is that Paul should prefer smoking (ψ_1 or ψ_2) to non-smoking ($\neg\psi$) and that my TORT mechanism captures this. Thus, according to Egan, ruling out the irrational options is all it takes to be a genuine ratifiability account. Since TORT does exactly that, the objection that TORT is not a genuine ratifiability account is not warranted.¹¹

2.2 So can TORT also account for other examples?

A possible objection to TORT could be that it is tailor-made to the specific example of the Three-Option Smoking Lesion. To refute this objection I provide two other examples where TORT is relevant.

Firstly, TORT is also applicable to the first example of the Psychopath Button since it has the same relevant elements as LRT. Recall that the result of step one is indecisive since both pressing and not pressing are not first-order (or second-order) ratifiable. We skip step two since there are no first-order ratifiable options. In step three we decide to apply EDT, and so

conditional on the fact that if Paul presses he has a great chance of being a psychopath, Paul does not press the button. And that again seems to be rational.

Secondly, TORT is also applicable in another recent example of 'Picking the Box'. Arntzenius presented Picking the Box in the paper of Gustafsson (2011: 149) in order to show how LRT can be refuted:

Paul is confronted with three boxes A, B, and C. However, Paul can only choose A or B in this first scenario. A perfect predictor has filled the boxes with money. If the predictor predicted that Paul will take A then he filled the boxes as follows: 2 euros in A, 1 euro in B, and nothing in C. If he predicted Paul will take B then he filled the boxes as follows: 4 euros in A, 3 euros in B, and nothing in C.

See the corresponding Table 2 below.

Choosing A	Choosing B
$VAL(A) = 2$	$VAL(A) = 4$
$VAL(B) = 1$	$VAL(B) = 3$
$VAL(C) = 0$	$VAL(C) = 0$

Table 2: Picking the box

It is rational for Paul to choose option B: he gets 3 euros instead of 2 euros, which is what he would get were he to choose option A. According to LRT, A is the only ratifiable option, and thus LRT would recommend the irrational option, namely A. Option B would be unrati-fiable because you would be better off choosing option A, on the supposition that B is chosen. However, TORT would get it right. Choosing option A is first-order ratifiable. And option B is second-order ratifiable since the first-order ratifiable alternative ($VAL(A) = 2$ euros) does not exceed the value of option B ($VAL(B) = 3$ euros). Now both option A and B are ratifiable, so accordingly Paul should choose the option with the highest VAL_{EDT} , namely option B.

3. Conclusion

In this essay I have argued for a transformed version of the Lexical Ratificationism Theorem (LRT): the Two-Order Ratificationism Theorem (TORT). We have seen that in order to recommend rational decisions, versions of Ratifiability are often added to Causal Decision Theory (CDT). According to proponents of CDT, imposing a ratifiability requirement will help us to save CDT. The ratifiability requirement teaches us that it becomes rational to perform an action *A* if and only if *A* is ratifiable. Two of such accounts are present in the current literature, the (original) Ratification Theorem (RT) and LRT. However, Egan and Gupta (Egan, 2007) came up with decisive counterexamples which showed that ratifiability accounts of CDT sometimes recommend the irrational action. Yet, standard Evidential Decision Theory (EDT) will also not endorse the rational action. This is bad news to decision theorists. Therefore, I introduce a first and second-order form of ratifiability to CDT. TORT can accommodate the intuitively compelling counterexamples to CDT that have recently been articulated by Egan (2007). By incorporating a second order, TORT is able to assess whether the first-order ratifiability account is the rational action. With the introduction of TORT, I have proved that there is a ratificationist account that is able to deliver the right rational choice in the sort of three-option cases that Gupta has pointed out.

The generalizability of my proposed TORT is yet to be tested. Above all, we should try to find counterexamples that prove my TORT to recommend the irrational decision. Or, we should try to find counterexamples that show TORT does not recommend the rational decision with the highest expected value for the agent.

Acknowledgements

I would like to thank dr. H.C.K. Heilmann for his insightful comments on earlier versions of this paper. He has given me in-depth feedback and remarkable guidance on my philosophical writing along the two-year masters program. Furthermore, thanks to the journal's anonymous reviewers, I was able to substantially improve and refine my work.

Nora Neuteboom (1989) holds her BSc degree in Business Economics from the University of Amsterdam in 2010. She obtained her MSc in Economics, Markets & Policy at the Erasmus University in 2012. She wrote her thesis at the econometrics department on the Influence of Economic Activity on the Onset of Civil War. In September 2012 she started the Research Master programme at the Erasmus Institute for Philosophy and Economics (EiPE). She is currently finishing her thesis on Metaphysics of Free Will and thereafter will do an internship at the Dutch Permanent mission at the United Nations in New York.

'Would You Press a Button that Kills All Psychopaths?' was written for the master's course 'Decision and Game Theory' with dr. H.C.K. Heilmann.

Notes

1. We do not need to worry about which of the many possible interpretations of utility we should endorse here. For our present purposes, this question does not matter since it is clear from the examples and counterexamples which action or option is preferred.
2. The term 'ratifiable' resulted from Jeffrey's idea that the agent can ratify a decision once it has been made. That is, his chosen action should have a maximal expected utility on the assumption that this is the action he is going to choose.
3. The term ratificationism was actually introduced by Jeffrey (1983: 19): 'Ratificationism requires performance of the chosen act, *A*, to have at least as high an estimated desirability as any of the alternative performances on the hypothesis that one's final decision will be to perform *A*'. But, to avoid confusion or different interpretations of the notion of ratificationism, I will only focus on the axiomatic notation of ratifiability conveyed by Egan.
4. This example shows us that imposing a ratifiability requirement will not help us to save CDT. It also shows that fans of EDT should take no comfort in the difficulties of this particular example against CDT. What we have here is definitely not an argument for a return to Evidential Decision Theory, since there are enough examples where EDT cannot be saved by RT either (Egan, 2007: 109-112).
5. One could argue that it is always better to have no guideline, than to have a wrong guideline. At least RT principle that counts all unratifiable actions as irrational will not deliver the bad recommendation that we got from the original version of CDT. But according to Egan (2007: 108) this does not do enough. He argues for 'completeness' where the correct theory of rational decision will endorse the rational action. Obviously, RT does not fulfill this requirement.
6. Notice that LRT takes the VAL_{EDT} instead of the value of VAL_{CDT} , which is used in the original RT.

7. One may wonder why this is called Lexical Ratificationism. As far as I know, this refers to the principle according to which entries in a dictionary are ordered, that is, the order depends on the first letter unless these are the same in which case it is the second which decides, and so on. I think this resembles the method where ratifiable actions are always to be preferred over unratifiable ones, but, within the categories, the action with greater VAL_{EDT} is to be preferred.

8. In the world of the Smoking Lesion, smoking is strongly correlated with lung cancer, but this correlation is understood to be the result of a common cause. There is a genetic lesion that tends to cause both smoking and cancer. The Smoking Lesion is an often used counterexample to both EDT and CDT.

9. I have not included a decision table for the Three-Option Smoking Lesion since such a table seems to be only more confusing. (Besides, a lot of spare paper was spilled finding out the right decision table). Instead, I think, a value table is more profitable for understanding the counterexample, so see table 1.

10. I have to admit that I do not find Gupta's counterexample very compelling. It seems peculiar that you would be better off smoking cigars when you are actually smoking cigarettes, and the other way around. I think such an odd thought would not be rational in the first place. However, this is not the focus of my paper. What is important is that the option not to smoke is irrational and should never be endorsed by any decision theorem.

11. Or, as an attentive reader might point out, to make sure that no options – even not option $\neg\psi$ – are ratifiable. He might say that if we un-ratify option $\neg\psi$, then we are at the same desirable point where we can treat all choices alike, and let the value according to EDT decide. But this is a mistake. Remember that all ratification mechanisms advance CDT. So, if there are no ratifiable options on the table, we have to make a decision according to CDT. CDT would tell us not to smoke since non-smokers tend to be best off. And so, CDT does not take into account the conditionality that, if we are to decide between option ψ_1 or ψ_2 , we are best off smoking.

12. Note that TORT is principally similar to LRT, except for the fact that it incorporates two orders of ratifiability.

13. Egan (2007: 108-109) gives two constraints for the adequacy of theory of rational decisions: soundness and completeness. A theory is sound if, when it's irrational to perform X, the correct theory of rational decision will not endorse doing X. A theory is complete if it's rational to perform X, the correct theory of rational decision will endorse doing X. My TORT mechanism satisfies both requirements. TORT advises Paul to smoke (which is a rational decision) and does not advise Paul not to smoke (which is irrational in the case of Paul to do).

References

- Egan, A. (2007) 'Some Counterexamples to Causal Decision Theory'. In: *The Philosophical Review* 116(1), 93-114.
- Gustafsson, J.E. (2011) 'A Note in Defence of Ratificationism'. In: *Erkenntnis* 75(1), 147-150.
- Jeffrey, R.C. (1983) *The Logic of Decision* (2nd ed.). Chicago: University of Chicago Press.
- Joyce, J.M. (1999) *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.
- Nozick, R. (1969) 'Newcomb's Problem and Two Principles of Choice'. In: Nicholas Rescher (ed.) *Essays in Honor of Carl G. Hempel*, 114–146. Dordrecht: Reidel.



Wittgenstein vanuit continentaal fenomenologisch perspectief

Myrthe van Nus

'Denk niet, maar kijk!'

Ludwig Wittgenstein (1953: 74)

Wittgenstein lezen vanuit een continentaal fenomenologisch perspectief zal bij meer dan een enkele lezer verbazing wekken. Het overgrote deel van de Wittgensteininterpretaties wortelt immers in de analytische filosofie waarin Wittgensteins werk wordt geduid in relatie tot werk van analytische filosofen als Frege, Russell, Moore, Carnap, Quine, Austin en Ryle (Kahane, Kanterian & Kuusela, 2007: 13). Naast deze analytische 'mainstreaminterpretatie' zijn verbanden tussen Wittgenstein en de continentale filosofie - weliswaar op kleinere schaal - echter ook onderzocht (Ibid.: 13). Men kan dan denken aan overeenkomsten tussen het werk van Wittgenstein en dat van Heidegger, Husserl, Merleau-Ponty, Nietzsche en Derrida of bijvoorbeeld onderzoek naar Wittgensteins gebruik van de term fenomenologie.¹ Deze interpretaties hebben gemeenschappelijk dat zij duidelijk maken dat er een bepaalde verwantschap bestaat tussen Wittgenstein en de continentale fenomenologie, maar komen niet verder dan het vergelijken en beschrijven van begrippen of de ontwikkeling van het gebruik van deze begrippen.

Dit leidt ertoe dat we niet weten of we Wittgenstein als verwant met de fenomenologische beweging kunnen beschouwen. Toch is het van belang dit te weten, omdat een positief antwoord op deze vraag impliceert dat we Wittgensteins gedachten met een alleen op de analytische filosofie georiënteerde interpretatie tekort zouden doen. Verder onderzoek zou dan gewenst zijn. De inzet van dit paper is dan ook om een eerste antwoord te formuleren op de vraag of we Wittgenstein nu wel of

niet tot de continentaal fenomenologische beweging kunnen rekenen en welke methode geschikt zou zijn om deze vraag te beantwoorden. Een tweede belangrijk doel van dit paper is aan te tonen dat het bestaande onderzoek naar Wittgenstein en de fenomenologische beweging ons niet verder helpt bij het beantwoorden van deze vraag. Dit is ook meteen de rechtvaardiging voor mijn eigen onderzoek naar een methode die deze vraag wel van een antwoord kan voorzien.

Na een korte kenschets van de continentaal fenomenologische beweging, laat ik aan de hand van een toepassing van de methodische criteria van Spiegelberg zien dat we Wittgenstein tot de beweging van de fenomenologie kunnen rekenen. Hierna bespreek en evalueer ik twee andere onderzoeklijnen: een lijn die Wittgenstein begripsmatig met de fenomenologen Husserl en Heidegger vergelijkt en een lijn die de ontwikkeling van de term 'fenomenologie' binnen het denken van Wittgenstein schetst. Deze twee onderzoeklijnen brengen ons niet veel verder, zo zal blijken. Daarom pleit ik in de conclusie voor een uitgebreider onderzoek naar de relatie tussen Wittgenstein en de fenomenologische beweging aan de hand van de methodische criteria van Spiegelberg.

1. De continentale fenomenologie

Wat wordt in dit paper nu verstaan onder de continentaal fenomenologische beweging? Een aanwijzing hiervoor vinden we in het *Jahrbuch für Philosophie und phänomenologische Forschung* uit 1913, een werk dat onder andere door de fenomenologen Husserl, Scheler en Pfänder werd samengesteld. Spiegelberg citeert hen:

‘It is not a system that the editors [Husserl, Geiger, Pfänder, Reinach & Scheler] share. What unites them is the common conviction that it is only by a return to the primary sources of direct intuition and to insights into essential structures derived from them that we shall be able to put to use the great traditions of philosophy with their concepts and problems; only thus shall we be in a position to clarify such concepts intuitively, to restate the problems on an intuitive basis, and thus, eventually, to solve them, at least in principle.’ (Spiegelberg, 1960a: 5)

Edmund Husserl is de grondlegger van deze beweging, maar ook de hermeneutisch fenomenoloog Martin Heidegger kunnen we volgens Spiegelberg tot deze beweging rekenen (Ibid.: 5, Spiegelberg, 1960b: 695).² Ik schrijf nadrukkelijk steeds ‘beweging’ en niet bijvoorbeeld ‘school’, ‘programma’ of ‘onderzoek’ omdat de term ‘beweging’ een belangrijke karakteristiek is van deze filosofie. De term ‘beweging’ slaat op het gegeven dat de fenomenologen vertrekken vanuit een gemeenschappelijk punt, maar dat hun filosofieën zich vervolgens in verschillende richtingen ontwikkelen en niet noodzakelijk bij hetzelfde eindpunt uitkomen (Spiegelberg, 1960a: 2).

Het gemeenschappelijk vertrekpunt bestaat uit een bepaalde houding van de fenomenologen: het beschrijvende onderzoek naar de fenomenen, zowel objectief als subjectief, in hun volle breedte en diepte (Ibid.: 2). Hun trouw aan de fenomenen is de attitude die de fenomenologen met elkaar delen, en deze attitude wordt vervolgens door hen uitgewerkt in methoden die verschillen per fenomenoloog. Bovendien is de fenomenologie gebruikt voor zeer verschillende doeleinden; denk aan de sociologie, psychologie en pedagogiek (Spiegelberg, 1960b: 653).

2. Wittgenstein *methodisch* vergelijken met de fenomenologische beweging

Hoe komen we nu vanuit de karakterisering van deze beweging bij een antwoord op de vraag of we Wittgenstein hiertoe kunnen rekenen? Door de kern van de fenomenologische methode te analyseren en deze te vergelijken met het denken van Wittgenstein. Spiegelberg heeft gezocht naar de kern van de fenomenologische methode door de gemeenschappelijke

aspecten te selecteren en de variabele aspecten in bepaalde mate te negeren (Ibid.: 653). Deze kern heeft hij uitgewerkt in methodische stappen waarvan de latere stappen meestal, maar niet noodzakelijk, op de eerdere stappen volgen (Ibid.: 655). De reden waarom hij zich op de methode richt is dat er binnen de fenomenologische beweging nauwelijks onenigheid bestaat over de opvatting dat de methode de karakteristieke kern van de fenomenologische beweging is (Ibid.: 655). De methodische stappen die Spiegelberg uitwerkt zal ik hieronder weergeven. Niet alle stappen van deze methode worden overigens door alle fenomenologen onderschreven: de eerste drie wel, de overige stappen alleen door een kleinere groep (Ibid.: 659). Sommige stappen bestaan uit een aantal onderdelen en sommige overlappen in bepaalde mate.

- a) Onderzoeken van specifieke fenomenen
- b) Onderzoeken van algemene essenties
- c) Vatten van essentiële relaties
- d) Wijzen van verschijning bekijken
- e) Onderzoeken hoe het fenomeen in het bewustzijn wordt geconstitueerd
- f) Staken van het geloof in de existentie van het fenomeen, oftewel de fenomenologische reductie toepassen
- g) Interpreteren van de betekenis van het fenomeen, het hermeneutisch aspect van de methode (Ibid.: 659).³

Hieronder zal ik aan de hand van Spiegelbergs methodische criteria een eerste aanzet geven tot het onderzoeken van de vraag of we Wittgenstein tot de fenomenologische beweging kunnen rekenen. Daarbij baseer ik me grotendeels op een later werk van Wittgenstein, de *Filosofische Onderzoekingen*.

2.1 Sub a: Onderzoeken van specifieke fenomenen

Deze stap bestaat uit meerdere onderdelen die in elkaar overlopen en daarom niet goed van elkaar te onderscheiden zijn. Hieronder vallen het onbevangen tegemoet proberen te treden van de fenomenen, het vatten

van hun onderscheidende constituerende aspecten en hun relaties met verwante fenomenen alsmede het beschrijven van het fenomeen en haar eigenschappen aan de hand van een classificatie (Ibid.: 659-673). De classificatie heeft als doel het fenomeen te positioneren ten opzichte van bestaande classificaties, de selectie om ons te beperken tot de beslissende karakteristieken van het fenomeen.

Wittgenstein probeert fenomenen onbevangen en objectief tegemoet te treden. Dit blijkt bijvoorbeeld in de *Tractatus Logico-Philosophicus* (hierna: *TLP*).⁴ De eerste stelling en substellingen van de *TLP* kunnen we lezen als een poging van Wittgenstein om puur te beschrijven wat zich aan hem toont: ‘De wereld is alles wat het geval is’ (Wittgenstein, 1961: 5). Wittgenstein probeert hier een zo objectief mogelijke uitspraak over de wereld te doen. Ook uit de *Filosofische onderzoekingen* spreekt deze attitude:

‘Alle verklaring moet terzijde worden geschoven, en alleen beschrijving moet ervoor in de plaats komen.’ (Wittgenstein, 1953: 94)

En:

‘Denk niet, maar kijk!’ (Ibid.: 74)

Dit laatste lezen we wanneer Wittgenstein ons attent maakt op de gelijkenissen en verschillen tussen spelen.

Wittgenstein probeert de eigenschappen van het fenomeen ‘spel’ te beschrijven. Dit blijkt lastig: verschillende soorten spelen hebben niet per se dezelfde karakteristieken. De één gaat om winst en verlies, de ander om behendigheid, weer een ander om het hebben van geluk en sommige om een combinatie van één of meer van deze eigenschappen. Uiteindelijk schrijft hij dat de spelen vanwege hun onderlinge verwantschap en verschillen het beste te karakteriseren zijn met de term ‘familiegelijkenissen’ (Ibid.: 74-75).

De term ‘familie’ duidt aan dat er een onderscheid is met andere families, maar ook dat de leden van een familie onderling verwant zijn. Precies zoals dat met bijvoorbeeld de getallenfamilie en de spelfamilie

het geval is: een spel uit de spelfamilie heeft gelijkenissen met de andere spelen uit deze familie, maar niet met getallen uit de getallenfamilie; deze hebben wel weer gelijkenissen binnen hun *eigen* familie.

We zien hier duidelijk dat Wittgenstein probeert om met een onbevangen attitude de fenomenen die hij onderzoekt tegemoet te treden. Wittgenstein toont bovendien oog te hebben voor zowel de onderscheidende aspecten van fenomenen als de gelijkenissen met andere verwante fenomenen.

2.1 Sub b: Onderzoeken van algemene essenties

Dit is het onderzoeken van voorbeelden van een essentie. Deze voorbeelden zijn gevallen die zich op een bepaalde manier verhouden tot de algemene essentie die deze voorbeelden kenmerkt (Spiegelberg, 1960b: 676-679).

Wittgenstein probeert in zijn werk de algemene essentie van het fenomeen ‘denken’ te vatten. Via een soort vrije associatie beschrijft hij welke begrippen te maken kunnen hebben met denken. Enkele voorbeelden:

- ‘Wanneer ons door een gedachte plotseling iets duidelijk wordt, wanneer we met andere woorden iets begrijpen, heeft dit te maken met denken’ (Wittgenstein, 1953:170). Begrijpen staat in nauw verband met het fenomeen denken.
- ‘Spreken en schrijven kunnen essentieel zijn voor denken, maar dit hoeft niet per se’ (Ibid.). Wittgenstein suggereert in dit verband dat we gewoonlijk nadenken tijdens spreken en schrijven, maar bespreekt ook een voorbeeld waaruit blijkt dat er niet altijd sprake is van denken wanneer men spreekt (Ibid.: 176-177).
- ‘Soms is het onduidelijk of er sprake is van denken, bijvoorbeeld wanneer we de juiste uitdrukking voor onze gedachten proberen te vinden. De uitdrukking kan opeens komen, zonder dat we werkelijk denken, maar ook kunnen we een beeld proberen te beschrijven dat we met de uitdrukking associëren, en in dit geval denken we waarschijnlijk wel.’ (Ibid.: 174-175)

Deze voorbeelden maken duidelijk dat Wittgenstein zoekt naar algemene essenties. Daarmee past zijn methode ook binnen de tweede stap van de fenomenologische beweging. Hij onderzoekt mogelijke verschijningsvormen van het fenomeen ‘denken’ om te weten te komen hoe deze zich verhouden tot de algemene essentie van denken.

2.1 Sub c: Vatten van essentiële relaties

Hier gaat het om het ontdekken van essentiële relaties (Spiegelberg, 1960b: 680-684). We kunnen twee situaties onderscheiden: de situatie waarin het gaat om interne relaties van een essentie (waarin we ons af moeten vragen of de componenten wel of niet essentieel zijn) en de situatie waarin het gaat over essentiële relaties tussen verschillende essenties. Zoals Wittgenstein de essentie van het fenomeen denken probeert te vatten, probeert hij ook de essentiële interne relaties weer te geven:

- ‘Denken drukt zich uit via de taal, dit is een essentiële relatie’ (Wittgenstein, 1953: 173). Wittgenstein noemt de taal in dit verband het voertuig van het denken.
- ‘Ook is denken menselijk, een machine denkt niet’ (Ibid.: 182). Menselijkheid als eigenschap van het subject dat denkt is zo bezien een essentiële relatie van denken.
- ‘Mogelijk, maar niet essentieel is twijfelen of je denkt of hebt gedacht’ (Ibid.: 173). Twijfel is op die manier een mogelijke relatie van denken.

Deze voorbeelden maken duidelijk dat Wittgenstein essentiële relaties van de fenomenen die hij beschrijft, probeert te vatten.

2.1 Sub d: Wijze van verschijning bekijken

Hiermee doelt Spiegelberg op het onderzoeken van de manier waarop een fenomeen aan ons verschijnt. Zo moet men letten op de kant van waaruit we een bepaald object waarnemen. Dit wil zeggen: de wijze waarop het fenomeen ‘gekleurd’ - want vanuit een bepaald perspectief - aan ons

verschijnt en de helderheid en onderscheidenheid van het object dat we waarnemen (Spiegelberg, 1960b: 684-687). Deze stap van de fenomenologische methode heeft ten doel dat we ons bewust worden van de gekleurdheid van onze waarnemingen.

Ook Wittgenstein wijst ons in zijn werk meerdere keren op de gekleurdheid van onze waarnemingen. Allereerst kunnen we dan denken aan de konijn-eendfiguur uit de *Filosofische onderzoekingen* (Wittgenstein, 1953: 135-136). Het kijken naar deze figuur impliceert een bewustwording van ons - natuurlijke - perspectivisme. De kant van waaruit we de figuur waarnemen is namelijk letterlijk bepalend voor wat we zien in de figuur. Monk meent bovendien dat het doel van Wittgensteins filosofische methodologie ligt in het kunnen veranderen van het perspectief van waaruit we naar de dingen kijken (1990: 508). We kunnen bijvoorbeeld een wiskundig bewijs als een serie proposities zien, maar ook als een afbeelding.

Ook in eerder werk van Wittgenstein - de *TLP* - vinden we aanwijzingen dat Wittgenstein zich terdege bewust was van de gekleurdheid van onze waarnemingen:

‘The world of the happy man is a different one from that of the unhappy man.’ (Wittgenstein, 1960: 87)

Iemand die gelukkig is, zal de wereld en alles erin heel anders bekijken dan iemand die ongelukkig is. Iedere waarneming heeft zo gezien iets interpretatiefs, en is in elk geval nooit geheel objectief. We interpreteren de wereld altijd vanuit een bepaalde gestemdheid.

Zelfs in zijn vroege werk houdt Wittgenstein zich dus al bezig met de wijze waarop fenomenen aan ons verschijnen en deze houding is in zijn latere werk nog sterker aanwezig.

2.1 Sub e: Onderzoeken hoe het fenomeen in het bewustzijn wordt geconstitueerd

Deze stap van de fenomenologische methode laat zich het duidelijkst uitleggen aan de hand van voorbeelden. Spiegelberg noemt de oriëntatie van iemand in een voor hem nieuwe stad. De stad krijgt langzaam vorm

in zijn bewustzijn (Spiegelberg, 1960b: 688-690). Zijn percepties vallen in een patroon en vormen relaties met elkaar. Wittgenstein schrijft iets vergelijkbaars over het leren van woorden:

‘Wanneer iemand bijvoorbeeld het woord ‘kubus’ tegen mij zegt, weet ik wat het betekent. Maar kan mij het hele gebruik van het woord voor de geest staan, als ik het zo *begrijp*?’ (Wittgenstein, 1953: 103)

Dit kunnen we als volgt interpreteren. Iemand die een woord kent qua naam, maar niet precies weet in welke contexten het allemaal gebruikt kan worden, is net als de toerist die de naam kent van een stad, maar de weg nog niet goed weet. Zoals de stad ons met de tijd steeds vertrouwer wordt, zal een woord dit ook worden.

Wittgenstein onderzoekt inderdaad hoe fenomenen in ons bewustzijn worden geconstitueerd.

2.1 Sub f: Staken van het geloof in de (non-)existentie van het fenomeen, oftewel de fenomenologische reductie toepassen

De reductie helpt ons om verschillende zintuiglijke waarnemingen zo gelijk mogelijk te behandelen (Spiegelberg, 1960b: 690-694). De fenomenologische reductie of *epoché* impliceert dat de dingen in de acten van het bewustzijn worden geconstitueerd en alleen daar werkelijk zijn wat ze zijn (Prins, 2007: 344). De fenomenologische reductie leidt er met andere woorden toe dat de werkelijkheid in haar eigenlijke gedaante aan ons kan verschijnen. Dit laatste is geenszins vanzelfsprekend. Tegenover de fenomenologische instelling staat namelijk de zogenaamde natuurlijke instelling die gegevens uit de diverse empirische wetenschappen gebruikt. Met deze natuurlijke instelling blijven volgens Husserl bepaalde aspecten van de dingen voor ons verborgen (Ibid.: 344). Om deze reden introduceert Husserl de fenomenologische attitude. In Husserls systeem van de wetenschappen is er dan ook een fundamenteel onderscheid. Enerzijds zijn er de *eidetische* wetenschappen (die betrekking hebben op het *wezen* / de essenties van de dingen (*eidōs*) en onafhankelijk van de empirische ervaring zekere kennis geven) en anderzijds de empirische wetenschappen (Prins, 1990b: 6).

Wittgenstein meent dat de taal van de wetenschap niet kan worden gebruikt om fenomenen te beschrijven, hij schuift deze aan de kant (Gier, 1981: 104). Bovendien wil hij - net als de fenomenologen - niet verklaren, maar ‘alleen’ beschrijven.⁵ Het gaat in Wittgensteins denken niet om het verkrijgen van empirische kennis over fenomenen, of het gebruiken van deze soort kennis om het *wezen* van het fenomeen te vatten, maar om het vatten van het *wezen* van de fenomenen via fenomenologisch onderzoek naar hun essenties en essentiële relaties. Bepaalde eigenschappen van fenomenen die via een empirische benadering niet blootgelegd kunnen worden, probeert Wittgenstein via zijn fenomenologische insteek te vatten.

Dit doet hij in de *Filosofische onderzoekingen* ook veelvuldig, bijvoorbeeld voor het fenomeen ‘taalspel’ (Wittgenstein, 1953: 73 e.v.), waarvan Wittgenstein het *wezen* probeert te vatten, maar ook voor het *wezen* van de taal (Ibid.: 89), van een omgrenzing (Ibid.: 91), van een (privé)beleving (Ibid.: 123-124, 158), van rekenen (Ibid.: 184 e.v.), van verwachtingen (Ibid.: 232 e.v.) en nog vele andere fenomenen. Soms maakt zijn bespreking van een fenomeen onderdeel uit van de definiëring van een ander fenomeen of blijken fenomenen met elkaar verwant te zijn.

We kunnen dan ook stellen dat Wittgenstein bij het beschrijven van bepaalde fenomenen de fenomenologische reductie toepast.

Op basis van de hierboven besproken voorbeelden uit het werk van Wittgenstein en de toepassing van de criteria van Spiegelberg op deze voorbeelden kunnen we voorzichtig concluderen dat de methode waarmee Wittgenstein zijn denken uitdrukt verwantschap lijkt te vertonen met die van de fenomenologische beweging.

2.2 Evaluatie

Wat hier niet mag ontbreken is een korte reflectie op de waarde van de criteria van Spiegelberg. Spiegelberg wordt gezien als *de* toonaangevende historicus van de fenomenologie (McKenna, 1983: 266). Maar wat moeten we bijvoorbeeld denken van Spiegelbergs opmerking dat niet alle methodische stappen door de toonaangevende fenomenologen onderschreven worden? Mijns

inziens vormt dit geen cruciaal probleem voor de criteria. Het punt is namelijk dat deze criteria alleen opgesteld konden worden door in bepaalde mate te generaliseren. Dit betekent dat sommige details, verschillen en nuancerings verloren gaan en sommige fenomenologen hun posities niet meer in de criteria herkennen. Echter, zonder de criteria zouden we alleen denkers één op één met elkaar kunnen vergelijken. We zouden dan alleen bundels van uiteenlopende criteria hebben, terwijl nu de hoofdlijnen van de fenomenologische beweging in de criteria getoond worden. We zullen later bovendien zien dat de meerwaarde van Spiegelbergs criteria er nu juist in bestaat dat het in bepaalde mate generalisaties zijn.

Nu kan men ook tegen de *selectie* van Spiegelberg argumenteren en bijvoorbeeld opwerpen dat ook Kant en Hegel tot de fenomenologische beweging behoren en hun werk niet binnen Spiegelbergs criteria valt. Spiegelberg zelf meent dat wanneer hij de grenzen veel minder strak zou hebben getrokken, ze *nog* minder specifiek en inhoudelijk zouden zijn geworden (1960b: 654). Het bereiken van uitputtende en definitieve resultaten was dan ook niet de insteek van Spiegelbergs werk.

3. Wittgenstein *begripsmatig* vergeleken met Husserl en Heidegger

Een deel van het reeds bestaand onderzoek naar Wittgenstein en de fenomenologische beweging vergelijkt Wittgenstein *begripsmatig* één op één met bepaalde fenomenologen. Dit soort onderzoek heeft dus de tegenovergestelde methodologische insteek van Spiegelbergs onderzoek, maar blijkt niet tot voldoende conclusies te kunnen komen.

Waar gaat het dan nu precies mis in dit soort onderzoek? Ik bespreek een aantal illustraties daarvan om dit te tonen. Methodische elementen komen in dit soort onderzoek als losse begrippen aan bod, en niet als fasen van de fenomenologische methode. Ik heb gekozen voor Husserl en Heidegger omdat deze fenomenologen passen binnen de fenomenologische context die ik in de eerste paragraaf heb geïntroduceerd.

3.1 Husserl en Wittgenstein

Gier is één van de auteurs die deze *begripsmatige* vergelijkingen toepast. De fenomenologische reductie behandelt Gier bijvoorbeeld wel (Gier, 1981: 104), maar als geïsoleerd element en niet als een onderdeel dat onlosmakelijk verbonden is met bepaalde andere elementen van de fenomenologische methode. Het grootste deel van Giers analyse bestaat echter uit een-op-eenvergelijkingen tussen de denkers.

Bijvoorbeeld over de data die zij gebruiken. De enige legitieme data voor zowel Wittgenstein als Husserl zijn de fenomenen van de onmiddellijke ervaring (Ibid.: 105). Het doel van de filosofie van Wittgenstein is om een *Darstellung* van de onmiddellijke ervaring te geven. Een *Darstellung* is een fenomenologische presentatie en datgene wat wordt gepresenteerd zijn geen zintuiglijke waarnemingen, maar fenomenen. Het gaat om een vorm van ervaring, en niet om een *feit* van ervaring (Ibid.: 81). Vergelijkbaar met Husserl meent Wittgenstein dat het binnen de filosofie belangrijk is om de essentie van datgene wat *dargestellt* is te vatten: 'In philosophy, it is important 'to get a full-blown pictorial *Darstellung* of our grammar [...]' (Wittgenstein, 1953: 165, Gier, 1981: 82). En over grammatica gesproken: een andere belangrijke overeenkomst tussen Husserl en Wittgenstein is gelegen in hun schrijven over grammatica. Beide menen dat er een filosofische grammatica bestaat die fundamenteeler is dan de traditionele logica (Gier, 1981: 99).

Maar er zijn ook verschillen tussen Husserl en Wittgenstein waarneembaar; sommige ontspringen zelfs aan wat eerst een overeenkomst leek. Qingxiong (2006) schrijft hierover. Zowel Husserl als Wittgenstein meenden dat er een vorm van ervaring is die voorafgaat aan alle ervaringsinhoud en die daarom *a priori*⁶ kan worden genoemd (Ibid.: 129). Volgens Qingxiong staat de term '*a priori*' bij Wittgenstein voor de interne structuur van de realiteit, die zichzelf toont in de taal, maar die niet via deze taal kan worden beschreven (Ibid.: 129). Husserl hanteert ook begrippen die op dit soort kennis betrekking hebben: *eidōs* en *eidetische* relaties.⁷ Echter, Husserl verschilt met Wittgenstein van mening over de eigenschappen van dit soort kennis (Ibid.: 131). Kortgezegd meent Husserl dat *a-priori-kennis* kan worden gevat via *eidetische* intuïtie (in plaats van door observatie of logische verificatie) terwijl Wittgenstein

meent dat is gebaseerd op regels van de syntax (Ibid.: 133). Overigens is dit ook weer een illustratie van hoe de vroege Wittgenstein van de *TLP* over dit onderwerp denkt.

3.2 Heidegger en Wittgenstein

Gier onderzoekt ook overeenkomsten tussen Wittgenstein en Heidegger. Zowel de filosofie van Heidegger als die van Wittgenstein blijkt geen theorie over sociologische of psychologische feiten te zijn, maar een descriptieve fenomenologie van levensvormen (Gier 1981, 44). Ook zijn beide filosofen zich bewust van de beperkingen van de taal en aarzelen zij om te zeggen wat in essentie onzegbaar is (Ibid.: 217). Daarnaast menen beiden dat een object en zijn Zijn ontologisch verschillend (different) zijn, maar desalniettemin onscheidbaar (Ibid.: 45). De crux is dat Heidegger schrijft dat het Zijn, als de voorwaarde voor de mogelijkheid van objecten, niet zelf een object kan zijn (ontologische differentie). En Wittgenstein schrijft in de *TLP*:

‘If I know an object I also know all its possible occurrences in states of affairs (=Being). Every one of those possibilities must be part of the nature of the object.’ (Wittgenstein, 1961: 6)

Mulhall op zijn beurt maakt een vergelijking tussen Heidegger en Wittgenstein op het gebied van Heideggers begrippen *Vorhanden* en *Zuhanden* (Mulhall, 1990). Het punt van Heidegger is dat objecten altijd aan ons verschijnen binnen een bepaalde concrete oriëntatie (Prins, 1990a: 4). Hieraan koppelt Heidegger de term *Zuhanden*. Wanneer de objecten deze concrete oriëntatie echter ontberen, wanneer ze *alleen* objecten zijn, zijn ze slechts *Vorhanden*. Het punt is dat bij Heidegger de structuren die de filosofie dient te verduidelijken gevonden worden in de dagelijkse ervaringen van het *Dasein*⁸ en dat bij Wittgenstein de nadruk ligt op de grammaticale structuren van de taal die duidelijk worden in het gewone - dagelijkse - taalgebruik (Mulhall, 1990: 120). Beide structuren verschijnen dus binnen een bepaalde, concrete oriëntatie. De taal is in het denken van Wittgenstein met andere woorden *Zuhanden*.

3.3 Evaluatie

Deze in steek die begripsmatige vergelijkingen maakt, is problematisch omdat de fenomenologie juist gekarakteriseerd wordt door haar methode. Een goede vergelijking moet daarom rekening houden met de fenomenologische methode. Ook wanneer de conclusie zou zijn dat de onderzochte begrippen allemaal overeen zouden komen, zou dit dus in het geheel niet betekenen dat we Wittgenstein tot de fenomenologische beweging zouden kunnen rekenen. We kennen na dit soort vergelijkingen namelijk enkel overeenkomsten en verschillen tussen Wittgenstein, Heidegger en Husserl op *begripsniveau*. Bovendien is dit een *beperkte* vergelijking, we zien steeds Wittgenstein *ten opzichte* van Heidegger en Husserl. Zolang we op deze manier - vanuit dit perspectief - naar Wittgenstein kijken, zullen we niet te weten komen of de filosofie van Wittgenstein werkelijk fenomenologisch is in de zin dat zijn attitude en methode passen binnen de fenomenologische beweging. We weten alleen dat Wittgenstein op bepaalde punten overeenkomsten vertoont met Husserl en Heidegger, en op andere punten van hen verschilt.

We moeten bedenken dat niemand ooit de wereld om zich heen volledig neutraal kan bekijken. Heidegger heeft ons erop gewezen dat onze interpretatie van de fenomenen om ons heen steeds wortelt in een bepaalde belangstelling (*Vorhabe*), een perspectief van waaruit het fenomeen wordt bekeken met het oog op een bepaald mogelijke uitleg (*Vorsicht*) en mede gevormd wordt door de begrippen waarin wij ons uitdrukken (*Vorgriff*) (Heidegger, 2009: §150). Men bekijkt en beoordeelt de wereld altijd al vanuit een bepaald perspectief. Dit perspectief vormt als het ware een bril tussen ons en de werkelijkheid, waardoor we bepaalde aspecten vergroot zien, maar andere minder goed of helemaal niet meer. Voor het Wittgensteinonderzoek impliceert dit dus dat we ons moeten realiseren dat we met het vergelijken van Wittgenstein met Husserl en Heidegger ook bepaalde aspecten kunnen missen. Dit punt klemt extra omdat Wittgenstein ons met de konijn-eendfiguur nu juist zelf op dit perspectivisme heeft gewezen.⁹

Stel bovendien dat we fundamentele verschillen zouden vinden tussen Wittgenstein en Husserl en/of Heidegger. Volgens Munson is dit

aannemelijk.¹⁰ We zouden dan in de verleiding kunnen komen hieraan de conclusie te verbinden dat Wittgenstein toch niet in de beweging van de fenomenologie thuishoort, terwijl het enige wat deze informatie ons werkelijk zegt is dat Wittgenstein van *hen* verschilt.¹¹ Immers, verschillen zeggen niet per definitie veel. Ook tussen de fenomenologie van Husserl en die van Heidegger bestaan bijvoorbeeld aanmerkelijke verschillen en ook hen rekenen we beiden tot de fenomenologische stroming. Ook impliceren verschillen tussen Wittgenstein en continentale denkers niet dat Wittgenstein toch verder van deze denkers afstaat dan van de analytisch georiënteerde filosofen. Immers, ook in het werk van analytisch georiënteerde denkers zijn namelijk genoeg verschillen met het denken van Wittgenstein te vinden.¹²

Daarnaast heeft het denken van Husserl en Heidegger net als dat van Wittgenstein een significante ontwikkeling doorgemaakt. Vergelijkingen tussen begrippen van deze drie denkers zijn bijna per definitie een vergelijking tussen bepaalde intellectuele perioden van deze denkers. Maar doordat de begrippen in de besproken onderzoeken als het ware geabstraheerd worden van hun normale context, is het risico aanwezig dat men niet altijd duidelijk voor ogen houdt welke periode van een bepaalde filosoof in de vergelijking wordt betrokken. Hoewel het werk van bijvoorbeeld Gier zeker nuttig is, slaat het op dit punt nogal eens deels de plank mis door wel te vermelden dat de vergelijking er één met - bijvoorbeeld - het latere werk van Wittgenstein betreft, maar niet te vermelden of het gaat om vroeger of later werk van Husserl dan wel Heidegger.¹³ Dit maakt de vergelijking onzorgvuldig en onvolledig van aard. Doordat Gier hierover geen informatie geeft, kan de lezer dit aspect onmogelijk zelf beoordelen.

Uit bovenstaande blijkt dat we de vraag of een bepaalde denker een fenomenoloog is niet kunnen beantwoorden door zijn begrippen te vergelijken met die van fenomenologen. Bovendien is er niet één bepaalde denker die als ijkpunt voor de continentaal fenomenologie kan worden genomen, waardoor de vergelijkingen altijd een-op-een-vergelijkingen zijn, en niet representatief zijn voor de fenomenologische *beweging*.

4. Wittgenstein over de *term* fenomenologie

Een belangrijke 'speler' in het onderzoek naar het gebruik van de term fenomenologie in Wittgensteins werk is Noë. Hieronder illustreer ik hoe Noë dit onderzoek aanpakt, zodat ik kan vaststellen wat dit onderzoek ons leert over de verhouding tussen Wittgenstein en de fenomenologische beweging. Dit doe ik aan de hand van een selectie van drie zaken die Noë beschrijft: wat Wittgenstein verstaat onder de termen 'fenomenologie' en 'fenomenologische taal' en waarom Wittgenstein op een bepaald moment zijn opvattingen over de fenomenologische taal wijzigde.

Wittgenstein, zo schrijft Noë, kwam tot de fenomenologische insteek omdat zijn poging om de epistemologie in zijn filosofie te incorporeren - zoals Bertrand Russell dit voor zijn eigen filosofie had gedaan - niet was geslaagd. Maar wat verstaat Wittgenstein dan precies onder fenomenologie? Fenomenologie is de naam die Wittgenstein toekent aan het onderzoek naar de aard van fenomenen (Noë, 1994: 8). Fenomeen en symbool moeten hetzelfde bereik aan mogelijkheden hebben. Zijn fenomenologie draait om het mogelijke, als tegenovergestelde van het actuele of waarschijnlijke (Ibid.: 8). Dit resulteert in een onderzoek naar de essentie.

Een voorbeeld kan dit verduidelijken. Bij een fenomenologisch onderzoek naar kleuren gaat het erom een symbolisme te construeren dat de kleuren helder representeert, en daartoe moet er sprake zijn van een perfecte representatie van dat wat wordt gerepresenteerd (Ibid.: 9).

Een tweede belangrijke term in dit kader is 'fenomenologische taal'. Voor Wittgenstein ligt het doel van deze taal in een perfecte representatie van dat wat wordt gerepresenteerd (Ibid.: 10). Neem kleuren; in dit geval staat de fenomenologische taal alleen een notitie toe waarin wat mogelijk is gerepresenteerd kan worden (geel of rood bijvoorbeeld) en waarin wat onmogelijk is (rood-groen bijvoorbeeld) uitgesloten wordt (Ibid.: 10).

Maar waar komt deze noodzaak van een fenomenologische taal vandaan? Die ligt in het gegeven dat volgens Wittgenstein de voorgestelde fenomenologische taal veel duidelijker en minder misleidend is dan onze gewone (fysische) taal (Ibid.: 10). Om dit te kunnen begrijpen moeten we fenomenologie en natuurkunde tegenover elkaar zetten. De natuurkunde maakt gebruik van hypothesen en hypothetische objecten in haar

verklaringen, de fenomenologie doet dit niet. Hoewel Wittgenstein erkent dat de gewone fysische taal ook haar voordelen heeft (ze kan eenvoudig beschrijvingen geven van zaken, terwijl een fenomenologische beschrijving enorm complex is), acht hij deze taal toch ongeschikt voor de representatie van de onmiddellijke ervaring (Ibid.: 11). De gewone fysische taal is 'slechts' geschikt voor het bespreken van zaken uit het fysische domein. Beide talen bieden ons dus weliswaar middelen om fenomenen te beschrijven, maar bij de gewone fysische taal ontstaat er door de beschrijving filosofische verwarring (Ibid.: 11-12).

Rond 1929 veranderde Wittgenstein zijn opvattingen over de fenomenologische taal. Noë meent dat deze verandering te maken heeft met Wittgensteins verwerping van het filosofisch project van de *TLP*: het creëren van een logisch transparante representatie van datgene wat verward wordt uitgedrukt in onze gewone taal (Ibid.: 17). In plaats daarvan moet de filosofie zich volgens Wittgenstein bezighouden met het zorgvuldig onderzoeken en vergelijken van verschillende methoden van representatie (Ibid.: 18). Hij heeft dus niet langer een voorkeur voor een meer correcte methode; verschillende methoden van representatie zijn geschikt voor verschillende doeleinden (Ibid.: 18-19).

Wittgensteins opvatting over de aard van de fenomenologie verandert: fenomenologie is nu gelijk aan grammatica. Dat wil zeggen dat fenomenologisch onderzoek niet meer is dan een onderzoek naar datgene wat zinvol is om te zeggen (Ibid.: 19). Het gaat wel nog steeds om mogelijkheden. Fenomenologisch onderzoek wordt een grammaticaal onderzoek van de woorden die gebruikt worden om de onmiddellijke ervaring te beschrijven.

4.1 Evaluatie

Nu we het wisselende fenomenologiebegrip van Wittgenstein hebben besproken kunnen we concluderen dat dit ons 'slechts' informatie oplevert over wat Wittgenstein verstaat onder de term 'fenomenologie' en over hoe deze term zich heeft ontwikkeld. De vraag of we Wittgenstein tot de fenomenologische beweging kunnen rekenen wordt door dit soort onderzoek echter niet van een antwoord voorzien. Dit komt doordat dit

soort onderzoek het denken van Wittgenstein in het geheel niet koppelt aan de methode van de fenomenologische beweging.

Uiteraard impliceert Wittgensteins gebruik van de term 'fenomenologie' op zichzelf geenszins dat hij als fenomenoloog kan worden gezien. Dit laatste weet ook Spiegelberg (Spiegelberg, 1960a: 7). Immers, anders zou een willekeur aan filosofen tot de fenomenologie kunnen worden gerekend, alleen maar omdat ze de term een paar keer hebben gebruikt. Dit zegt niet alleen niets over de fenomenologische aard van hun werk, maar heeft hetzelfde effect op de fenomenologische beweging: deze zou haar onderscheidende karakter verliezen wanneer het noemen van de term fenomenologie voldoende is om tot de beweging gerekend te worden.

5. Conclusie

Het is gebleken dat het vergelijkend toepassen van de fenomenologische methode van Spiegelberg op later werk van Wittgenstein een eerste aanwijzing oplevert dat we Wittgenstein kunnen rekenen tot de fenomenologische beweging. Deze insteek biedt een nieuw perspectief op Wittgensteins werk, omdat deze niet slechts begrippen vergelijkt of hun ontwikkeling in de tijd onderzoekt, hetgeen ons geen antwoord heeft opgeleverd op de vraag of we Wittgenstein tot de fenomenologische beweging kunnen rekenen, maar Wittgensteins denken direct binnen de fenomenologische beweging en haar methode situeert. Het belang van het onderzoek dat ten grondslag ligt aan dit paper is dat een interpretatie van Wittgenstein die uitsluitend in de analytische filosofie wortelt zijn denken niet volledig kan vatten en recht kan doen. Verder onderzoek is uiteraard noodzakelijk om dit volledig en met zekerheid te kunnen vaststellen.

Myrthe van Nus LL.M. B.A. (1986) is masterstudent Wijsgerige Antropologie aan de Faculteit der Wijsbegeerte van de EUR en student-assistent bij meerdere leerstoelgroepen aan dezelfde faculteit. Eerder rondde zij de master Rechtsgeleerdheid aan de EUR cum laude af. Momenteel is zij bezig haar wijsgerige masterscriptie over pessimisme af te ronden. Hierin laat zij zien waarom pessimistische filosofen als Schopenhauer tekortschieten en dat Nietzsches idee van het pessimisme uit kracht de voorkeur verdient.

Noot van de redactie

Omdat dit essay is geschreven door een oud-redactielid van het Erasmus Student Journal of Philosophy, is een extra zorgvuldige review-procedure gevolgd. Zie voor meer informatie <http://www.eur.nl/fw/esjp/inzendingen>.

Noten

1. Zie bijvoorbeeld: Gier (1981) respectievelijk Noë (1994).
2. In 1927 werd *Zijn en Tijd* van Martin Heidegger ook gepubliceerd in dit Jahrbuch.
3. Omdat Spiegelberg schrijft dat hij sterk aarzelt om deze stap tot de fenomenologische methode te rekenen (Spiegelberg, 1960b: 694), pas ik deze hieronder niet toe op het denken van Wittgenstein.
4. Dit werk wordt gerekend tot Wittgensteins vroege periode.
5. Dit onderscheid tussen de empirische wetenschappen en wetenschappen van essenties is ook voor een ander aspect van belang. Sommigen menen dat de filosofie van Wittgenstein geen fenomenologie is omdat een transcendentale methode zou ontbreken en zijn filosofie daarom slechts het predicaat fenomenalisme verdient (een fenomenalist meent dat alles wat bestaat slechts zintuiglijke indrukken zijn van objecten, en niet de werkelijke objecten zelf). Echter, een fenomenalist zou het onderscheid dat Wittgenstein maakt tussen de twee soorten wetenschappen verwerpen (Gier, 1981: 95). Wittgenstein schrijft hier zelf ook over: 'What sort of investigation is ours? Do I investigate the cases which I bring up according to their probability or factuality? No, I bring up only that which is possible by giving grammatical examples' (Gier: 1981: 95, Wittgenstein, 1933: 425). Daarnaast is de filosofie van Wittgenstein volgens Gier wel degelijk transcendentaal van karakter nu Wittgenstein geïnteresseerd is in de logica (essentie) van fenomenen (Gier, 1981: 94).
6. Kennis die voorafgaat aan de ervaring noemt men *a-priori-kennis*.
7. Eidetische relaties zijn essentiële relaties die het transcendentale bewustzijn constitueert binnen de verschillende sferen van het eidetische. Door deze relaties worden verschillende dingen naar hun essentie, hun wezen, met elkaar in verband gebracht.
8. De term *Dasein* duidt niet op een bepaald vermogen of eigenschap, maar op het *er-zijn* van de mens, op de gesitueerde *voltrekking* van het menselijk bestaan, dat telkens door een zeker verstaan van zichzelf en van de dingen is doortrokken (Prins, 2007: 156). Tevens articuleert de term het *er-zijn* van de mens: de fundamentele openheid waarin het *Dasein* de wereld ontmoet. De wereld is niet de som der dingen, maar het onpeilbaar geheel van betekenis- en verwijzingsamenhangen, waarin het *Dasein* zich voltrekt (Ibid.: 156). Heidegger schrijft daarom dat het 'wezen' van het *Dasein* in zijn existentie ligt (Heidegger, 2009: §42).
9. Zie hiervoor: Wittgenstein (1953: 135-136) en Monk (1990: 508).

10. Munson benadrukt het eigen karakter van de fenomenologie van Wittgenstein en de afstand met de fenomenologie van Husserl, maar ook met die van Heidegger (1962: 46).

11. Zie in dit verband ook: Park (1998: 206-207). Park waarschuwt in zijn werk *Phenomenological aspects of Wittgenstein's philosophy* ook voor mogelijk overhaaste conclusies naar aanleiding van vergelijkingen als die tussen Wittgenstein en Husserl. De reden dat ik Park in dit paper niet verder bespreek is dat Park de oorsprong van de fenomenologische invloed op Wittgenstein niet zoekt in de continentale filosofie, maar in een hele andere stroming, namelijk die van theoretisch natuurkundigen als Boltzmann, Hertz en Mach (Ibid.: viii, 207). Het gaat dan om invloed vanuit de fenomenologische natuurkunde, die ook van invloed is geweest op het denken van Husserl. Een bespreking van de ideeën van Boltzmann, Hertz en Mach gaat het bestek van dit paper te buiten.

12. Bijvoorbeeld tussen Wittgenstein en Frege (zie hiervoor: Gier, 1981: 205-206) of tussen Wittgenstein en de logisch positivisten (zie hiervoor: Qingxiong, 2006: 133). Zie voor verschillen tussen Wittgenstein en Frege ook: Wittgenstein (1953: 137).

13. Zie bijvoorbeeld: Gier (1981: 107), waar hij het concept transcendentaal ego van Husserl vergelijkt met het latere werk van Wittgenstein. Dat het transcendentaal ego pas werd ontwikkeld door de latere Husserl wordt in het geheel niet genoemd. Dit is echter wel van belang om de term te plaatsen, nu tegenover het transcendentaal ego de eerder door Husserl ontwikkelde term empirisch ego staat. Gier noemt dit laatste begrip helemaal niet.

Literatuurlijst

- Gier, N.F. (1981) *Wittgenstein and Phenomenology: A Comparative Study of the Later Wittgenstein, Husserl, Heidegger and Merleau-Ponty*. Albany: State University of New York Press.
- Heidegger, M. (2009) *Zijn en Tijd*. Nijmegen: SUN.
- Kahane, G., Kanterian, E. & Kuusela, O. (eds.) (2007) *Wittgenstein and His Interpreters: Essays in Memory of Gordon Baker*. Oxford: Blackwell Publishing.
- McKenna, W.R. (1983) 'The Context of the Phenomenological Movement'. In: *Journal of the History of Philosophy* 21(2), 266-267.
- Monk, R. (1990) *Ludwig Wittgenstein: The Duty of a Genius*. Londen: Vintage.
- Mulhall, S. (1990) *On Being in the World*. Londen: Routledge.
- Munson, T.N. (1962) 'Wittgenstein's Phenomenology'. In: *Philosophy and Phenomenological Research* 23(1), 37-50.

- Noë, R.A. (1994) 'Wittgenstein, Phenomenology and What it Makes Sense to Say'. In: *Philosophy and Phenomenological Research* LIV(1), 1-42.
- Park, B.C. (1998) *Phenomenological Aspects of Wittgenstein's Philosophy*. Dordrecht: Kluwer.
- Prins, A.W. (2007) *Uit verveling*. Kampen: Klement.
- Prins, A.W. (1990a) 'Existenciale analyse van het Dasein: Heidegger'. In: *Syllabus Fenomenologie en hermeneutiek*. Rotterdam: Erasmus Universiteit.
- Prins, A.W. (1990b) 'Grondslagen van de fenomenologie: Husserl'. In: *Syllabus Fenomenologie en hermeneutiek*. Rotterdam: Erasmus Universiteit.
- Qingxiong, Z. (2006) 'Wittgenstein's Reconsideration of the Transcendental Problem – With some Remarks on the Relation Between Wittgenstein's 'Phenomenology' and Husserl's Phenomenology'. In: *Philosophical Research* 10, 68-76.
- Spiegelberg, H. (1960a) *The Phenomenological Movement. A Historical Introduction. Volume one*. Den Haag: Martinus Nijhoff.
- Spiegelberg, H. (1960b) *The Phenomenological Movement. A Historical Introduction. Volume two*. Den Haag: Martinus Nijhoff.
- Wittgenstein, L. (1961) *Tractatus Logico-Philosophicus*. Londen: Routledge.
- Wittgenstein, L. (1953) *Filosofische onderzoekingen*. Amsterdam: Boom.
- Wittgenstein, L. (1933) *The Big Typescript* (niet gepubliceerd).

