



# meta( $\varphi$ )

## Reduktion, Repression, Realisation

**Functional Reduction as an Alternative to Nagelian Reduction**

• Nina Peier

**A Problem with Functional Models of Theory Reduction**

• Manuel Steffen

**Sex Work is Work**

• Mo Brunold

**The Myth of Reverse Discrimination**

• Karen Poertzgen

**Wittgensteins Parallelen zu der buddhistischen Realitätsvorstellung**

• Omar Ibrahim

**Der Wille zur Wahrheit im Kontext der Kritik asketischer Ideale**

• Khalil Staubli

**Herausgeberin**

Fachschaft Philosophie der Universität Bern,  
Länggassstrasse 49a, 3012 Bern, Schweiz  
mit Unterstützung des Instituts für Philosophie der Universität Bern.

**Redaktion**

meta(φ)  
Fachschaft Philosophie, Universität Bern  
metaphi.philo@lists.unibe.ch

**Chefredaktion**

Timo Junger

**Lektorat**

Alexandra Birrer, Raffael Grolimund, Vera Moser, Maurizio Piu, Malin Plikat, Xenia Schmidli,  
Andrea Schnyder

**Produktion und Gestaltung**

Aleksandar Nikolić

**Coverbild**

Kathrin Stettler

**Druck**

Kromer Print AG Lenzburg

**ISSN**

Gedruckte Ausg.: meta(φ) ISSN:2297-9948  
Online-Ausg.: meta(φ) [Elektronische Ressource] ISSN:2297-9956

**Text-Einreichungen**

Einreichungen zur Veröffentlichung sind erwünscht an [metaphi.philo@lists.unibe.ch](mailto:metaphi.philo@lists.unibe.ch). Die Auswahl der Texte erfolgt themenunabhängig, jedoch wird philosophische Relevanz des Themas vorausgesetzt. Für jede Ausgabe akzeptieren wir bis zu zwei Einreichungen pro Person. Es besteht kein Anrecht auf Veröffentlichung, weder bei erfolgreicher Einreichung noch nach Anfrage durch ein Redaktions-Mitglied. In diesem Magazin veröffentlichte Texte gelten als wissenschaftliche Publikation und sind somit zitierfähig.

**Spenden**

Die Produktion des meta(φ) ist kostspielig. Sofern Ihnen das Journal gefällt, freut sich die Fachschaft Philosophie daher sehr über finanzielle Unterstützung. Spenden werden folgender Empfängerin zuteil.

Fachschaft Philosophie  
Universität Bern, 3012 Bern  
IBAN: CH53 0900 0000 3050 3703 2  
Zahlungsgrund: «metaphi»

# Erste Worte

Liebe Leser\*innen

Wenngleich philosophische Texte manchmal schwer verständlich und kompliziert anmuten, so ist die Philosophie doch meist darauf aus, einfache und schlichte Erklärungen oder Theorien zu entwickeln. In diesem Zusammenhang entstehen unter anderem Versuche, Komplexes auf Fundamentaleres zu reduzieren. Gleichwohl muss die Philosophie auch anerkennen, dass sich nicht alles beliebig reduzieren lässt. Unsere Mitmenschen, zum Beispiel, sollten wir nicht bedenkenlos auf einzelne Kategorien wie Herkunft, Geschlecht, Alter oder sexuelle Orientierung herunterbrechen. Solche Beschreibungen wären nicht nur eindimensional und inadäquat, sondern würden zusätzlich die Gefahr bergen, verschiedenste Formen von Unterdrückung hervorzubringen, zu perpetuieren und vermeintlich zu legitimieren. Philosophische Theorien, welche die menschliche Realität in ihrer Ganzheit fassen, sind nicht einfach zu realisieren - wenn überhaupt. Nichtsdestotrotz versucht das meta( $\varphi$ ) durch die Verbreitung sechs philosophischer Essays, einen kleinen Beitrag zum Verständnis unserer Realitäten zu leisten.

Den Auftakt zu dieser siebten Ausgabe namens *Reduktion, Repression, Realisation* macht Nina Peier mit ihrem Essay über zwei Ansätze der Theorienreduktion. Die Autorin argumentiert, dass das von Jaegwon Kim entwickelte funktionale Reduktionsmodell keine Vorteile gegenüber dem von Ernest Nagel vorgeschlagenen Ansatz mit sich bringt. Thematisch eng verwandt

ist Manuel Steffens Text, in welchem eine weitere Kritik am funktionalen Modell hervorgebracht wird. Unter anderem wird argumentiert, das Modell sei einem bösartigen Regress ausgesetzt. Im darauffolgenden Teil wendet sich das Journal zwei Texten zu, die wir unter dem Leitmotiv *Repression* zusammengestellt haben. Zuerst beschäftigt sich Mo Brunold mit Debra Satz' *Why Some Things Should Not Be for Sale* und argumentiert, Sexarbeit lasse sich nicht in der von Satz vorgesehenen Weise von anderen Arbeitsmärkten unterscheiden. Anschliessend untersucht Karen Poertzgen, ob *affirmative action* als Form von Diskriminierung beschrieben werden kann und spricht in diesem Kontext vom *Mythos der umgekehrten Diskriminierung*. Die letzten beiden Texte drehen sich um Realität und Wahrheit. Omar Ibrahim thematisiert die Unterscheidung zwischen *Welt* und *logische Form der Wirklichkeit* bei Ludwig Wittgenstein und stellt gleichzeitig Parallelen zur buddhistischen Realitätsvorstellung her. Im abschliessenden Text fragt sich Khalil Staubli, wieso sich Menschen eigentlich philosophische Fragen stellen. Diesem Rätsel geht er im Kontext dessen nach, was Friedrich Nietzsche *Wille zur Wahrheit* nennt.

Ich wünsche Ihnen eine inspirierende Lektüre, spannende Gedanken und gute Gesundheit.

Herzlichst, Timo Junger  
im Namen des Redaktionsteams



Alexandra Birrer



Raffael Grolimund



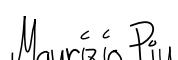
Timo Junger



Vera Moser



Aleksandar Nikolić



Maurizio Piu



Malin Plikat



Xenia Schmidli



Andrea Schnyder

“What are the words we do not yet have? What do you need to say? What are the tyrannies you swallow day by day and attempt to make your own, until you will sicken and die of them, still in silence?”

Audre Lorde  
Sister Outsider, 1984

“All that can be done is to plant the seeds of thought. Whether something vital will develop depends largely on the fertility on the human soil, though the quality of the intellectual seed must not be overlooked.”

Emma Goldman  
Anarchism and Other Essays, 1910

# Inhaltsverzeichnis

---

## Reduktion

Functional Reduction as an Alternative to Nagelian Reduction

- Nina Peier

6

A Problem with Functional Models of Theory Reduction

- Manuel Steffen

13

---

## Repression

Sex Work is Work

Arguments against viewing sex work as a morally distinctive labour market

- Mo Brunold

21

The Myth of Reverse Discrimination

Does Affirmative Action Discriminate Against the Majority?

- Karen Poertzgen

27

---

## Realisation

Wittgensteins Parallelen zu der buddhistischen Realitätsvorstellung

- Omar Ibrahim

36

Der Wille zur Wahrheit im Kontext der Kritik asketischer Ideale

Untersuchung der Bedeutung eines Begriffs in Nietzsches Genealogie der Moral

- Khalil Staubli

47

# Functional Reduction as an Alternative to Nagelian Reduction

## Introduction

In this essay, I am going to discuss the question whether the functional model of reduction, proposed by Jaegwon Kim (1998), is a more fruitful account of reduction than the traditional model proposed by Ernest Nagel (1961). Based on the argumentation of Ausonio Marras (2006), I am going to argue for the thesis that the functional model does not bring any merits compared to the Nagelian model.

I am going to proceed as follows: In the first section, I am going to reconstruct Kim's criticism of the Nagelian model of reduction. The most important point of this criticism is that Kim does not accept 'bridge laws' in reduction, because they remain unexplained and because they can only be formulated if all the vocabulary of the reducing and the reduced theory is known. After this, I am going to present his alternative model of reduction; functional reduction. According to Kim, functional reduction does not require bridge laws because the reduction process can be accomplished by the 'functionalization' of the reduced concept (Kim 2006, 552).

In the second section, I will discuss one of Ausonio Marras' objections to Kim's model. He argues that Kim does not succeed in avoiding bridge laws, once the process of functionalization has been clearly formulated. His thesis is that, in effect, functional reduction does not differ from Nagelian reduction. In addition to that, Marras returns to the Nagelian model, arguing that bridge laws are not as problematic for Nagelian reduction as Kim believes (see Marras 2006, 338).

I agree with Marras on the claim that functional reduction does require bridge laws. Kevin Morris, however, objects, that even if bridge laws are required, functional reduction does still differ from Nagelian reduction in an important way (Morris 2009, 653-654). He argues that functional reduction is a conjunction of Nagelian reduction and what he calls the 'functional exhaustion thesis'

(Morris 2009, 653). In the third section of this essay, I will reconstruct Morris' argument for the difference between the two models. After this, I am going to argue that the functional exhaustion thesis is not sufficient to elaborate a substantial difference between the two models of reduction.

## 1. Models of Reduction

The Nagelian model of reduction, which has been proposed by Ernest Nagel (1961), is probably the most popular and most discussed account to reduction. Jaegwon Kim (1998; 2005) criticizes this model, arguing that 'bridge laws', which are a necessary part of Nagelian reduction, are problematic. As an alternative to the Nagelian model of reduction, Kim proposes a functional model of reduction (Kim 2005, 101).

In this section, I will reconstruct Kim's arguments against the Nagelian model of reduction. In a second step, I am going to introduce Kim's model of functional reduction.

### 1.1 Nagelian Reduction and Its Problems

The Nagelian model of reduction provides an account to the reduction of theories (Nagel 1961, 336-397). It enables the reduction of higher-level theory to a lower-level theory, including their laws. Furthermore, it explains how deriving laws of the higher-level theory from laws of the lower-level theory is possible (see Nagel 1961, 338-366). As Kim formulates it, Nagelian reduction generally enables the reduction/derivation of predicates of one theory to another (Kim 1998, 90). In general, the vocabulary of two theories of which the higher-level one is to be reduced to the lower-level one is distinct. This is why predicates of the higher-level theory cannot be a priori connected to predicates of the lower-level theory without some connecting

principles. The connecting principles, which are usually referred to as ‘bridge laws’, are an important part of Nagel’s model. Bridge laws build reductive links between the vocabulary of a higher-level theory and the one of a lower-level theory<sup>1</sup> (Nagel 1961, 353–356; Kim 1998, 90). At this point it is to note that, whilst Nagel strictly speaks about the reduction of theories, Kim’s focus is on the reduction of properties included in such a theory (Morris 2009, 648). In the following, when I consider the reduction of properties, I refer to property predicates included in a theory. According to Kim, besides providing a reductive explanation, bridge laws enable the derivation of predicates of the higher-level theory from predicates of the lower-level theory (Kim 1998, 90).

Kim believes that bridge laws are “*at the heart of*” Nagelian reduction (Kim 1998, 90; 2006, 552). He considers bridge laws as a necessary condition for Nagelian reduction (Kim 1998, 91). Kim criticizes that bridge laws remain “*unexplained auxiliary premises*” (Kim 1998, 96). Since bridge laws can only be correctly formulated if all the vocabulary of both theories—the reducing one (lower-level theory) and the reduced one (higher-level theory)—is known, a lack of explanation occurs (Kim 1998, 95–96). Furthermore, there is no general way of detecting how the predicates of two theories must be related to each other in a reductive way (Kim 1998, 91). In summary, it is not possible to logically derive predicates of the higher-level theory from the vocabulary of the lower-level theory without using bridge laws. The missing explanation of how higher-level predicates can arise lawfully from lower-level predicates is called an ‘explanatory gap’ (Levine 1983, 354). Kim criticizes that the use of bridge laws does not fill the explanatory gap because, by using bridge laws, it is simply assumed what needs to be explained:

“In using the bridge laws as auxiliary premises of reductive derivations, the Nagelian reductionist is simply assuming exactly what needs to be derived and explained if we are to answer the explanatory questions raised by Huxley, James, and the emergentists—that is, if we are to close the explanatory gap or solve the hard problem of consciousness.”

(Kim 2005, 100)

In this quote, Kim mentions two problems connected to the explanatory gap which have to be solved in order to have a favorable model of reduction: first the hard problem of consciousness and second the problem of emergence. The hard problem of consciousness consists in the missing explanation of how conscious experience connects to the physical world. According to Kim, in order to solve the hard problem of consciousness, a model of reduction which is sensitive to mental properties, such as conscious experience, should be taken into consideration (Kim 2006, 554). The model should, furthermore, allow the possibility of emergent properties because the idea of emergent properties seems to be appealing for many philosophers (see Kim 2006, 547–548). If there are specified conditions that determine in which cases higher-level properties can be reduced to lower-level ones, and additionally, in which cases the former emerge from the latter, emergence could be explained (Kim 2006, 548). It is to note that Nagel himself does not deny the possibility of emergence (see Nagel 1961, 372). However, Kim criticizes that, as long as an explanatory gap remains, the required sensitivity to the hard problem of consciousness and the possibility of emergent properties cannot be explained by the Nagelian model of reduction.

According to Kim, if reduction is successful, it brings the merit of ‘true reductive explanation’ or ‘ontological simplification’ (or both). A model which would bring true reductive explanation would not face the problem of an explanatory gap as, following Kim, Nagel’s model does. Ontological simplification is fulfilled when fewer entities are needed in an ontology after the reduction than before. If mental properties were completely reducible to physical properties, there would be no need to postulate them as entities in the world (Kim 1998, 95–96).

In Kim’s view, due to the explanatory gap, a model of reduction which requires bridge laws enables neither true reductive explanations nor ontological simplifications (Kim 1998, 95–97; Marras 2006, 562). Thus, as long as the explanatory gap exists, Nagelian reduction is not successful. Kim believes that the only possibility to close the explanatory gap when keeping bridge laws would be to turn them into identities (Kim 1998, 98). Kim argues “*Identity takes away the logical space in which explanatory questions can be formulated*” (Kim 1998, 98). If bridge laws were identities, the explanatory gap would be closed, and the reduction would be successful.

However, turning bridge laws into identities is problematic due to the possibility of multiple realization (Kim 1992, 4–5). Very briefly, multiple realization means

---

<sup>1</sup> At this point it is to note that whilst Nagel mentions the reduction of theories, Kim is more interested in the reduction of properties which are included into a theory (Morris 2009: 648). When I mention the reduction of properties I refer to property predicates included in a theory.

that the same type of f.e. mental property can be realized by different types of physical properties. If we require bridge laws to be identities, multiple realization is not possible. The problem is located in the transitivity of the identity relation: The transitivity of identity implies that if a supervening property is identical with two different base properties, it follows that the two base properties must be identical as well. This is clearly not the case in cases of multiple realization. Hence, the identity requirement fails (see Kim 1992, 1-26).

To conclude, the only way in which bridge laws could provide true reductive explanations or ontological simplification is to turn bridge laws into identities. However, they cannot be identities because of the possibility of multiple realization. Thus, bridge laws provide neither true reductive explanations nor ontological simplification. Hence, a model of reduction which contains bridge laws—as the Nagelian model does—is not able to successfully explain the reduction of theories or properties. That is why Kim proposes an alternative model of reduction which he calls ‘functional reduction’.

## 1.2 The Alternative: Functional Reduction

In Kim’s alternative model of reduction, ‘reduction through functionalization’, or short, ‘functional reduction’, the notion of bridge laws is avoided (Kim 2006, 552). Briefly, functional reduction proceeds in two steps (see Kim 2005, 101-102): First, the higher-level properties, which should be reduced, are functionally conceptualized. This means that a functional definition must be found for the concepts of the higher-level properties. A functional definition identifies the causal specifications of a property. The complex of the causal specification is called ‘the functional role of the property’. If such a definition can be found, the property is ‘functionalizable’. Often, the properties, which are defined functionally, are called ‘functional properties’.

The second step of functional reduction consists in searching for physical realizers which ‘fill the role’ of the functional property (Kim 2005, 101). If either the concept of a higher-level property cannot be functionalized or no physical realizer fills the role of the functional property, the reduction is not successful (see Kim 2005, 101).

Kim notes that, when thinking about the functionalization of mental properties, we have to distinguish between three players; the brain state (bearer of the physical property), the mental state (mental property) and the behavior (function of the mental property). Whereas the relation between the mental state and the functional behavior is necessary, the relation between the physical realizer and the mental state is contingent (Kim 2006,

553). With this in mind, it is clear how functional reduction can deal with multiple realization: Due to the fact that the base property is variable, it is possible that multiple physical realizers fill the functional role. The derivation of the higher-level property from the lower-level property follows once we have identified the role that a physical property fills, because the relation between the functional property and the mental property is necessary.

To summarize, in contrast to the Nagelian model, the functional model of reduction does not postulate bridge laws. Instead, Kim claims that the logical derivation of higher-level properties is possible by functionalization. Since there are no additive unexplained premises, the functional model avoids the explanatory gap and succeeds in ontological simplification.

Another advantage of functional reduction compared to the Nagelian model is that it satisfies Kim’s demands for sensitivity towards emergent properties. Kim characterizes emergent properties as being supervenient, but irreducible (Kim 2006, 555-556). With the functional account of reduction, these conditions can be met.

## 2. Objections Against Functional Reduction

In his texts ‘Consciousness and Reduction’ (2005) and ‘Emergence and reduction: Reply to Kim’ (2006), Ausonio Marras challenges Kim’s model of reduction. His main thesis is that Kim fails in avoiding bridge laws. Due to this, Marras questions whether functional reduction is essentially different from Nagelian reduction. In Marras’ proposal for an explication of the notion ‘filling the role’, there is an analogy arising between the Nagelian model and the functional model. According to Marras, bridge laws are required in both theories. However, he does not think that whenever bridge laws are required, an explanatory gap exists. In contrary, he believes that there is a way in which bridge laws can be turned into identities. In this section, I am first going to explain the analogy Marras draws between the two models of reduction. Second, I am going to reconstruct Marras’ arguments of how bridge laws can be turned into identities and thus how they avoid the implication of an explanatory gap.

### 2.1 Explication of ‘Filling the Role’

Ausonio Marras (2005; 2006) argues for the strong thesis that functional reduction does not substantially differ from Nagelian reduction because it does not succeed in avoiding bridge laws. Marras criticizes that Kim does not give a sufficient explication of what it means

for a realizer to ‘fill the role’ of a functional property (Marras 2006). Marras claims that, once such a sufficient explication is articulated, the functional model of reduction does not substantially differ from the Nagelian model anymore (Marras 2005; 2006).

As explained in section 1, Kim developed the model of functional reduction in order to—at least to some extend—avoid the notion of bridge laws. Due to this, it is clear that it would be fatal for Kim’s model if Marras’ thesis proved to be true: If functional reduction required bridge laws, an explanatory gap, such as identified in Nagelian reduction, could not be avoided. If this was the case, functional reduction would neither succeed in giving reductive explanations nor in providing ontological simplification.

Marras proceeds in the following way: First, he examines how Kim explains what is meant by ‘filling the role’, i.e. what it takes for a physical property to realize a functional property (see Marras 2006, 346). In the functional model of reduction, a mental property  $M$  can be reduced to a physical property  $P$ , if and only if  $P$  meets  $M$ ’s causal specifications. According to Kim,  $M$ ’s causal specifications can be articulated by identifying the set of causal relations  $M$  enters with other properties in the same system  $S$  (see Kim 2005, 101). If the causal specifications for a mental property  $M$  have been identified, it is further called the functional property. For example, the causal specifications of ‘belief’ as a functional property would be determined by its relations to other beliefs, desires, etc. (Marras 2005, 345).

Second, Marras criticizes that this is not a sufficient explication of what it means to ‘fill the role’ of  $M$ . This is why he offers an improved explication of what it means for  $P$  to be a realizer of  $M$ : Suppose a functional property  $M$  is realized by a physical property  $P$  in a system of type  $S$ .  $M$  is a predicate of a higher-level theory  $T^*$  and  $P$  is a predicate of a lower-level theory  $T$ . Given this,  $M$  causally interrelates with properties in  $T^*$ , linked by laws  $L^*$  of  $T^*$ . To fill the role of  $M$ , i.e. to realize  $M$ ,  $P$  must causally interrelate with properties in  $T$  which are analogous to the properties  $M$  interrelates with in  $T^*$ . There have to be analogous laws  $L$  linking the property  $P$  with the causally interrelated properties of  $T$  (Marras 2005, 346).<sup>2</sup>

Third, Marras claims that the interrelating properties of  $T^*$  and  $T$  must be linked biconditionally. If this is the case,  $P$  fills the role of  $M$  in  $S$ . In other words, if there are biconditional laws between the properties of

$T$  and  $T^*$ , a functional property  $M$  can be realized by  $P$ . These biconditional laws are connecting principles which cannot be further explained. This formulation is very similar to what Nagel says about bridge laws (see Nagel 1961, 354). Thus, the analogy to bridge laws can be drawn. In effect, bridge laws are nothing but connecting principles between predicates of a higher-level theory and a lower-level theory. Marras concludes that functional reduction thus requires some sort of bridge laws as well. If bridge laws are required in functional reduction, an explanatory gap opens. This is, according to Marras, why functional reduction faces the same problems as Nagelian reduction (Marras 2005, 346-347).

Hence, offering a true ‘alternative’ to the Nagelian model of reduction seems to have failed. Kim’s aim was avoiding bridge laws in his functional account. If the two models work analogously, whenever Nagelian reduction requires bridge laws, so does functional reduction.

## 2.2 Bridge Laws as Theoretical Identities

Marras does not think that functional reduction is substantially different from Nagelian reduction. However, this does not imply that reduction does not work at all. Marras provides a solution to the problem of the identity of properties which Kim identifies in Nagel’s account. As mentioned in section 1, Kim takes bridge laws to be a necessary part of Nagelian reduction. In his view, the only way a reduction by means of bridge laws can be successful is when these laws are turned into identities. As I have argued in section 1.1, Kim considers this requirement to be problematic. Marras does not agree with Kim, offering a different interpretation of what it means for bridge laws to be turned into identities.

Marras argues that it is indeed possible to turn bridge laws into identities, or, as he calls it, that they can be ‘enhanced into identities’ (Marras 2005, 340). He argues that bridge laws used in reductions are not merely nomologically co-instantiated, but nomologically necessary. Nomological necessity of a bridge law means: Whenever a higher-lever or lower-level property occurs in the context of our laws of nature, the other property occurs as well. In other words, since it is impossible, within the context of natural laws, that one of the properties—connected to another property by a bridge law—occurs without the other one, the bridge law is nomologically necessary (Marras 2005, 340-341). Despite of this, Marras agrees with Kim that bridge laws are contingent in some way. He interprets this contingency as a merely logical contingency. Given

<sup>2</sup> Marras differentiates between the mental property and the functional property (see Marras 2005, 346). Due to reasons of simplicity, I abandon this difference for the purposes of this essay.

this, bridge laws can still be nomological necessary. In addition to that, Marras characterizes bridge laws as (nomologically) necessary *a posteriori* (Marras 2005, 341). In short, it means that the necessity of bridge laws can only be detected *a posteriori*, hence on an inductive basis. Marras infers from this that, on an inductive basis, biconditional bridge laws can be turned into identities. Marras calls those inductively identified identities ‘theoretical identities’ (Marras 2005, 340-341).

Whenever a bridge law is identified, it is on an inductive basis (Marras 2005, 341). Since bridge laws are in fact inductively derived principles, they do not remain unexplained auxiliary premises (see Marras 2005, 344). Given that bridge laws are theoretical identities, property identification is possible with the notion of bridge laws. Furthermore, Marras argues that once such theoretical identities are found, the explanatory gap does not exist anymore (Marras 2005, 566). What remains is what he calls a ‘conceptual gap’, which is not further problematic for the reduction process:

“A conceptual gap remains, and it can no more be closed [...]. But the existence of a conceptual gap does not entail the existence of an explanatory gap. For we can still get explanations as a result of theoretical property-identifications, even when these are inductively based of successful theory reductions.” (Marras 2006, 566)

### 3. The Merit of Functional Reduction

As shown in section 2.1, Marras defends the view that bridge laws can be part of a successful reduction. Furthermore, Marras does not think that functional reduction can avoid the notion of bridge laws. He infers from the latter that functional reduction does not essentially differ from Nagelian reduction. This consequence may be rejected. Kevin Morris (2009) has formulated a defense of functional reduction. He argues that, even if functional reduction requires bridge laws, it is still the better model of reduction than Nagelian reduction. Morris argues that functional reduction can be understood as Nagelian reduction plus the functional exhaustion thesis (see Morris 2009, 647-657). In this section, I am first going to examine what is meant by the functional exhaustion thesis and why Morris thinks that this is an important part of reduction. Second, I am going to defend Marras against the objection at hand: I am going to argue that the functional exhaustion thesis fails, once explicated, in drawing a sharp line between functional and Nagelian reduction.

#### 3.1 The Functional Exhaustion Thesis

Morris argues that functional reduction can be seen as the conjunction of Nagelian reduction and another condition he calls ‘functional exhaustion’ (Morris 2009, 650-651). One component of functional reduction, Morris agrees with Marras, is that the role of a mental property  $M$  has to be identified. This is also where the analogy between functional reduction and Nagelian reduction can be drawn. Morris reconstructs Marras’ thesis about the role of a mental property  $M$  as follows:

“[R] The role of  $M$  for structure type  $S$  is  $C$  according to theory  $T_m$  iff: There are properties  $M_1, M, \dots, M_n$  and laws  $L_{m1}, L_{m2}$  of  $T_m$  about  $S$  such that

[ $L_{m1}(M_1 \dots M \dots M_n), L_{m2}(M_1 \dots M \dots M_n)$ , etc], where  $L_m(M_1 \dots M \dots M_n)$  is to be read as:  
‘ $M$  interrelates with  $M_1 \dots M_n$  according to law  $L_m$  of  $T_m$ ’” (Morris 2009, 650)

In addition to [R], Morris formulates the functional exhaustion thesis, which he takes to be an additional part of the functional reduction model. The functional exhaustion thesis postulates that the functional role of a property is *exhaustively* characterized in a theory (Morris 2005, 469). Neither in the original formulation nor in Marras’ reading of the Nagelian reduction, the condition of functional exhaustion is required. Morris argues that, once we accept this condition, there is a substantial difference between Nagelian and functional reduction: The consequence of requiring functional exhaustion as a necessary condition for reduction is that some properties are reducible under the Nagelian account but not under the functional account of reduction. This holds even if functional reduction requires bridge laws (Morris 2009, 653).

If functional reduction is seen as [R] plus the functional exhaustion thesis, functional reduction is substantially different from Nagelian reduction, since the latter does not pay attention to functional exhaustion. One could argue that it serves as a more fine-grained account of reduction, since there are properties which are not reducible under the functional account of reduction but which are reducible under the Nagelian account.

#### 3.2 Implications of Functional Exhaustion

I agree with Morris that the functional exhaustion thesis is not explicitly mentioned in either Nagelian reduction or in Marras’ reading of functional reduction. However, I doubt that it is important for reduction at all. Morris does not give an exact explication of what

functional exhaustion means. Roughly, it means that in a system  $S$ , all the relevant causal relations which property  $M$  enters must be included in its functional conceptualization. If this is the case, the concept of  $M$  is functionally exhausted (Morris 2009, 653). What it means to identify all the relevant causal relations is quite important. If we do not know how to do that, reduction as Morris proposes cannot be accomplished. As I see it, there are two options:

The first option is that all possible causal relations which  $M$  is entering are part of the functional concept of  $M$ . If this is the case, a substantial difference to Nagelian reduction shows. In Marras' argumentation there is no consideration of all the possible causal relations of  $M$  but only the causally immediate. As Marras and I have argued, the latter necessarily have their counterparts in a lower-level theory such as it is in Nagelian reduction.

The identification of all possible properties is a strongly problematic requirement since it can only be fulfilled on an ontological level. However, it is epistemically impossible to give an exhaustive characterization of all possible relations  $M$  enters. If for no mental property all possible causal relations can be identified, no mental property can be functionalized and thus—no mental property can be reduced to a physical property. This is why I doubt that this reading of functional exhaustion guides to the model of reduction which Kim has in mind.

The second option is that all the relevant causal relations include the properties which are directly caused by  $M$ , or a cause of  $M$ ; in other words, all the immediate causal relations of

$M$ . However, this is the idea in Marras' reading of functional reduction as well. As Marras argues, we cannot find a difference between the models of reduction if we explicate the functionalization of mental properties. This is, because every property or law which stands in a causal relation to the functional property must have a counterpart in the lower-level theory. One could argue that the functionalization does help in finding the counterpart of  $M$  in the lower-level theory, but this would still not be a sufficient explanation of the difference between the two models of reduction.

One claim Morris makes is that there are properties which are not functionalizable, and thus irreducible under the functional account, whereas they are reducible under the Nagelian account (Morris 2009, 650–651). Morris himself does not give any examples for such cases. I suppose, what he has in mind are for example sensory states (*qualia*) or consciousness

itself. According to Kim, sensory states are not functionalizable and thus not reducible (Kim 2006, 553): “*I am with those who do not believe pain and other sensory states (“qualia”) can be given functional characterizations.*”

If we accept that under the functional account, sensory states and consciousness as a whole cannot be reduced, it still is not clear whether functional reduction has any merits compared to Nagelian reduction. In my view, it is unclear whether Nagelian reduction is able to reduce such properties from a higher-level theory to a lower-level theory. Hence, we cannot simply conclude that the striking difference between the two models lies in the ability/disability of reducing for instance sensory states to physical properties.

In summary, there is no striking difference between functional reduction and Nagelian reduction which has been identified so far. It thus remains an open question which account of reduction is more successful. What results from my analysis is, however, that functional reduction is not a simple alternative to the problematic Nagelian model. In contrary, it has to deal with the same explanatory questions connected to bridge laws. Moreover, it postulates even further premises by the notion of causal relations or even causal exhaustion which require further explanation.

## Conclusion

In this essay, I have discussed Jaegwon Kim's model of functional reduction which is supposed to serve as an alternative to Nagelian reduction. I have argued on the basis of Ausonio Marras' argumentation that, if sufficiently explicated, functional reduction does not succeed in avoiding connecting principles which are analogous to bridge laws. Avoiding the notion of bridge laws seems to be a major point in Kim's argumentation to illustrate the difference between the two models. Since he does not succeed in avoiding bridge laws, there is no big advantage of his model of reduction compared to the Nagelian in this. However, when accepting theoretical identities as Marras does, the use of bridge laws can be sufficiently explained and justified. It is thus not very problematic for a model of reduction. This led me to the question whether there is any other factor which shows the merit of functional reduction, compared to Nagelian reduction. For this reason, I have discussed an argument by Kevin Morris, which introduces the functional exhaustion thesis. In Morris' view, functional exhaustion is an important part of functional reduction. It is an addition to what Nagelian reduction does. Thus, functional reduction

is Nagelian reduction plus the functional exhaustion. Once we try to explicate what functional exhaustion means, it becomes unclear again whether there is a substantial difference to Nagelian reduction.

I do not propose to just reject functional reduction and return to Nagelian after this analysis. However, what I have shown is that there is no really good argument yet why we should prefer the functional model. To find such an argument, we would have to dig deeper into the purposes of scientific reduction, asking the question: What aims is an account of reduction supposed to serve?

## References

- Kim, Jaegwon. 1992. "Multiple Realization and the Metaphysics of Reduction." *Philosophy and Phenomenological Research* 52: 1-26.
- . 1998. *Mind in a physical world*. Cambridge University Press, MA: MIT Press.
- . 2005. *Physicalism, or Something Near Enough*. Princeton, NJ: Princeton University Press.
- . 2006. "Emergence: Core ideas and issues." *Synthese* 151: 547-559.
- Levine, Joseph. 1983. "Materialism and Qualia: The Explanatory Gap." *Pacific Philosophical Quarterly* 64: 354-361.
- Marras, Ausonio. 2005. "Consciousness and Reduction." *British Journal for the Philosophy of Science* 56: 335-361.
- Marras, Ausonio. 2006. "Emergence and reduction: Reply to Kim." *Synthese* 151: 561-569.
- Morris, Kevin. 2009. "Does Functional Reduction Need Bridge Laws? A Response to Marras." *British Journal for Philosophy of Science* 60: 647-657.
- Nagel, Ernest. 1961. *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York: Harcourt, Brace, and World, Inc.

---

**Nina Peier (24)** befindet sich am Ende ihres Bachelorstudiums in Philosophie und Sozialwissenschaften. In ihrem Studium befasst sie sich vorwiegend mit Fragen der Metaphysik, darüber hinaus interessiert sie sich für Wissenschaftsphilosophie, Epistemologie sowie für Methoden der Philosophie.

# A Problem with Functional Models of Theory Reduction

## Introduction

In past years, Michael Esfeld, Christian Sachse, and Patrice Soom (2007; 2012)<sup>1</sup> have proposed a new model of reduction, the declared goal of which is to provide an account of theory reduction that escapes the multiple realization argument—an argument often thought to *prima facie* refute any form of reductionism. The authors claim that they accomplish this goal by uniting the traditional bridge-law model of theory reduction due to Ernest Nagel (1961) with concepts borrowed from David Lewis's (1966; 1972; 1981) and Jaegwon Kim's (1998; 2005) models of functional reduction. Specifically, they hope to evade multiple realization by introducing the novel notion of a *functional subtype* (see Esfeld and Sachse 2007, 5-6; see Esfeld et al. 2012, 227-228). In accordance with this, they describe their model as “theory reduction by means of functional subtypes” (hereafter referred to as ‘subtype reduction’ or ‘SR’) (see Esfeld and Sachse, 2007).

My goal in this paper is to demonstrate that, although subtype reduction may in a certain sense resolve the issues it is intended to tackle, one major problem remains. This problem, I will argue, emerges in SR as soon as functional reduction is thought of not as an isolated instance of *ontological reduction*, but considered, as Esfeld et al.<sup>2</sup> do, in the context of epistemological *theory reduction*. Starting from this problem, my discussion will culminate in the contention that SR faces the following dilemma: Either it fails at its outset; or it remains a workable model—but only at the cost of its ability to provide reductive explanations.

In the context of philosophy of mind, for example, the latter horn of this dilemma would imply that SR

is left unable to close the ‘explanatory gap’ and thus fails to improve upon the various models it is intended to supersede. But the problem I will highlight also has a more general dimension: Since it affects functional subtypes but does not *arise* from them, it will extend beyond SR and trouble all similar cases of what I call *functional theory reduction*, where the functional model is integrated into, or read as, a model of theory reduction.

Before I go on to present my arguments in greater detail, I will elaborate on SR in the subsequent section, focusing especially on the way in which Esfeld et al. motivate the model and on their presentation of Nagelian and functional reduction. I will not spend much time with functional subtypes, as they are not as important for present purposes as the discussion that precedes them.

## 1. Subtype Reduction and Its Motivations

Esfeld and Sachse (2007) and Esfeld et al. (2012) contain very useful enumerations of the metaphysical and generally philosophical premises of subtype reduction. In the following, I briefly recount these premises before moving on to SR’s main motivations.

First, Esfeld et al. in both papers take a broadly realist stance towards theories, supposing that mature scientific theories contain both concepts which seize natural kinds and “laws or law-like generalizations, which capture salient causal connections in the world” (Esfeld et al. 2012, 217). Such theories may be ordered according to their breadth of domain of application, such that the objects of ‘higher-level’<sup>3</sup> theories also fall into the broader domains of more basic theories.

<sup>1</sup> Note that while Michael Esfeld and Christian Sachse were involved in both papers, Patrice Soom is only credited in the latter.

<sup>2</sup> Throughout this paper I will use “Esfeld et al.” to refer to Michael Esfeld and his co-authors. This may or may not include Patrice Soom, depending on the paper cited.

<sup>3</sup> The usage of the *levels-metaphor* has recently been criticised, for example in Heil (2016). For the purposes of this paper, I will nevertheless accept the metaphor and occasionally speak of ‘higher-level’ and ‘lower-level’ theories and properties.

Similarly, the properties represented by concepts in a higher-level theory strongly supervene on lower-level (and ultimately physical) properties, or on configurations of such properties (see Esfeld et al. 2012, 218). This does, however, not imply that lower-level properties are any more ‘real’ than higher-level or *special science* properties: All properties captured by mature theories are equally ‘real’ and exist in the same way (see Esfeld and Sachse 2007, 1).

Second, Esfeld and Sachse’s 2007 paper contains additional, specifically metaphysical premises. Among these is the acceptance of Kim’s *causal exclusion argument* and its consequences,<sup>4</sup> as well as a denial of the existence of property-types<sup>5</sup> (see Esfeld and Sachse 2007, 2). More importantly, however, Esfeld and Sachse in this paper also advance the following two theses, which are intended to characterize their reductionist position (*ibid.*, 1):

**Ontological Thesis:** All there is in the world are entities that are identical with fundamental physical tokens and their configurations.

**Epistemological Thesis:** It is in principle possible to reduce every true description of what there is in the world to a description constructed in terms of fundamental physics.

The two theses are then supplemented by the following, not entirely uncontroversial claim, which states that there is a two-way relation of entailment holding between them:

(C) Epistemological reduction and ontological reduction entail one another (*ibid.*, 2).

This claim does not appear in Esfeld et al. (2012), but since it assorts quite well with the realist stance towards theories described above, it should be safe to assume that Esfeld et al. are committed to it in both statements of SR. Indeed, upon closer inspection, it seems like (C) is in fact *central* to the project of SR. For it is apparent that if epistemological reduction (i.e. theory reduction) and ontological reduction entail one another, it turns out that a single model may account for both types of reduction. This, it seems to me, is exactly what Esfeld et al. intend SR to achieve in uniting Nagelian reduction with functional reduction. But even more, (C) conversely suggests that models of reduction originally designed to account

for only one of epistemological and ontological reduction will henceforth need to be reinterpreted as also accounting for the other.

I believe that these consequences of claim (C) are of great import, and I will return to consider them at the end of the paper. For now, they should serve as a backdrop to the introduction of SR and the emerging problems.

## 1.1 Nagelian Reduction and Multiple Realization

The main motivation for subtype reduction is the *multiple realization argument*, which is commonly raised in objections against Nagelian reduction.

Nagelian reduction is the traditional model of theory reduction, which states that a (usually upper-level) theory  $T_2$  reduces to a (usually lower-level) theory  $T_1$  if and only if the laws of  $T_2$  can be derived by logical or mathematical means from the laws of  $T_1$  (see Nagel 1961, 338). In most noteworthy cases of Nagelian reduction, however, the reduced and reducing theories concern vastly different properties and consequently, it is likely that a significant portion of the concepts distinctive of  $T_2$  will not be found in  $T_1$  and vice versa. In order for the reduction of  $T_2$  to  $T_1$  to nevertheless succeed, the Nagelian reductivist therefore needs to introduce bridge-laws: sentences which connect the concepts unique to  $T_2$  to those unique to  $T_1$  (see *ibid.*, 352-353).

Historically, there has been much debate about the form and status of bridge-laws—the two main points of contention being whether they are to be understood as biconditionals or as some other relation; and whether or not they are indeed laws, in the sense of being included in and extending  $T_1$ . Esfeld et al. commit themselves in the first matter to the more ‘traditional’ reading: Bridge-laws are to be understood as universally quantified biconditionals, which express (at least) nomic coextensions between  $T_2$ -concepts and  $T_1$ -concepts (see Esfeld et al. 2012, 219).<sup>6</sup> In the second matter, they follow Marras (2005) and other commentators in denying that bridge-laws need be laws in any strict sense. Rather, bridge-laws are auxiliary premises not integrated into  $T_1$ , but merely conjunctively added to it in order to facilitate the reduction of  $T_2$  (see *ibid.*).

In summary, Esfeld et al. take it that, in order to derive a law  $L$  containing a concept  $M$  unique to  $T_2$  from the

<sup>4</sup> See Kim 1998, 37-47 and 2005, 32-52 for thorough statements of the exclusion argument.

<sup>5</sup> In this paper I will remain, wherever possible, neutral on the issue of the existence of property-types.

<sup>6</sup> It is interesting to note that recent appraisals of Nagelian reduction point to Nagel’s own pluralism about bridge-laws and claim that on a faithful construal, they are neither biconditionals, nor susceptible to the multiple realization argument. See, e.g., Sarkar (2015). Due to a lack of space, I will not engage with these claims here and stick to Esfeld et al.’s reading.

laws of  $T_1$ , one requires auxiliary bridge-laws of the following form, connecting  $M$  to a suitable concept  $P$  in  $T_1$  (and thus also connecting all  $M$ -tokens to  $P$ -tokens):

$$\square_N \forall x (Mx \leftrightarrow Px)^7 \quad (1)$$

Unfortunately, however, the Nagelian account of reduction so construed faces a near insurmountable challenge—the multiple realization argument: Since the property types (and/or tokens) captured by the concepts of  $T_2$  strongly supervene on those captured by the concepts of  $T_1$ , multiple  $T_1$ -properties or property configurations may be eligible as supervenience bases for a single  $T_2$ -property. Thus,  $T_2$ -properties may be multiply realized by  $T_1$ -properties (or property configurations).

This thwarts reduction, for it denies the possibility of biconditionally linking the  $T_2$ - and  $T_1$ -concepts capturing the relevant properties. One can at most achieve universally quantified one-way conditionals running from each realizer-concept  $P_1 \dots P_n$  in  $T_1$  to the  $T_2$ -concept they realize:<sup>8</sup>

$$\square_N \forall x (Mx \leftarrow P_1 x) \wedge \dots \wedge \square_N \forall x (Mx \leftarrow P_n x) \quad (2)$$

Yet such one-way conditionals cannot facilitate the derivation of  $T_2$  from  $T_1$ . Thus, the argument from multiple realization spells doom for Nagelian reduction.

## 1.2 Functional Reduction and Elimination

Faced with the problem of multiple realization, Esfeld et al. look to *functional reduction* for a solution. As its name suggests, functional reduction is closely related to functionalist conceptions in philosophy of mind, which claim—broadly speaking—that mental states and properties are definable solely in terms of their functional (i.e. causal) roles.<sup>9</sup> Accordingly, functional reduction is at the outset a form of ontological *property reduction*, which seeks to reduce functionalizable properties represented by concepts of higher-level theories (e.g. mental properties) to properties represented by concepts of lower-level theories (e.g. physical properties). Functional reduction usually proceeds in three steps (see Kim 2005, 101–102; cf. Kim 1998, 98–

99; Esfeld and Sachse 2007, 3; Esfeld et al. 2012, 220):

**Step 1—Functionalization:** A property  $Q$  that is being targeted for reduction is given a functional definition in terms of a causal role  $H$ , which captures  $Q$ 's typical causes and effects.<sup>10</sup> This turns  $Q$  into a second-order property:

Having  $Q =_{\text{def.}} \text{Having a property which performs causal role } H$ .

**Step 2—Realization:** One then looks for some property (or configuration of properties)  $R$  which in fact performs the causal role  $H$  and thus *realizes*  $Q$ .

**Step 3—Integration:** Finally, one finds or constructs a theory which explains *how*  $R$ , the realizer of  $Q$ , brings about the effects that characterize  $H$ .

Reduction is then effected in the wake of these three steps by *identifying* the realizer property  $R$  with the functionalized target-property  $Q$  (see Kim 1998, 98–99).<sup>11</sup> This result agrees with Esfeld et al.'s ontological thesis: Successful ontological reduction is supposed to lead to the identification of higher-level entities with lower-level entities (see Esfeld and Sachse 2007, 1) and functional reduction provides such identifications. Yet functional reduction as stated here cannot evade the multiple realization argument. For it is apparent that many distinct lower-level properties may perform the same causal role  $H$  and realize the same functional property  $Q$ . In the context of theory reductions, functional reduction so far characterized therefore cannot provide the needed biconditionals and amounts to nothing more than the very same one-way conditionals that were already described in proposition (2). The traditional strategy for fortifying functional reduction against multiple realization is to construe the model as delivering *local reductions* (see Kim 2005, 25–26).<sup>12</sup> Esfeld et al., however, do not endorse such a

<sup>10</sup> Often, the typical causes of the reduced property are left out of the functional definition in favour of greater focus on the typical effects.

<sup>11</sup> Kim's argument has been criticised by Ausonio Marras (2000), who suggests that there are serious metaphysical problems with identifying second-order properties with their first-order realizers. Further criticism of various aspects of Kimian functional reduction may be found in Bregant (2006), Eronen (2011), Kistler (2005), and Marras (2005), among others.

<sup>12</sup> Since I do not have the space here to provide a thorough elabora-

7 The index “N” on the necessity operator is meant to indicate nomological necessity.

8 Unless one is willing to admit into the biconditional what Fodor calls a “heterogenous and unsystematic disjunction” of all possible realizers (Fodor 1974, 108).

9 For a short introduction to functionalism, see Block (1980).

construal. They argue that the local reduction of a multiply realizable property-type  $Q$  inevitably leads to the fragmentation and subsequent reductive *elimination* of whichever theory  $T_2$   $Q$  is represented by.<sup>13</sup>

If this is indeed the case, it immediately follows from claim (C) that the elimination of  $T_2$  entails the elimination of  $Q$ —which runs counter to Esfeld et al.’s central assumption that the special sciences describe real causal structures and concern genuinely existing properties. It thus seems like functional reduction is not fit to deal with multiple realization.

### 1.3 Functional Subtypes

How, then, are we to solve the problem of multiple realization? Esfeld et al.’s answer, as anticipated in the introduction, is that we are to solve it by introducing *functional subtypes*.

The notion of a functional subtype amends functional reduction by capitalizing on the fact that different realizers of a causal role  $H$  must, if they are truly to be *different*, also have differing causal dispositions (see Esfeld et al. 2012, 225). These differences in causal disposition must find expression in different ways of bringing about the effects that characterize  $H$ —and Esfeld et al. argue that such differences will be detectable not only in the reducing theory  $T_1$ , but also in the reduced theory  $T_2$ .

If this is correct, it allows for the construction of more finegrained causal roles  $H_1, \dots, H_n$  (i.e. functional subtypes), which account for the differences among  $H$ ’s realizers, but may be formulated using  $T_2$ -concepts exclusively (see *ibid.*, 226; cf. Esfeld and Sachse 2007, 5–7). This prevents the elimination of  $T_2$ , since it protects the theory from being split into multiple sub-theories, and, unlike local reduction, manages to uphold  $T_2$ ’s unique laws and law-like generalizations (see Esfeld et al. 2012, 228–229). Finally, introducing functional subtypes yields the needed biconditionals, which allow for the derivation of  $T_2$  from the laws of  $T_1$ .

Reduction via functional subtypes therefore seems to embody a promising solution to the problem of multiple realization, all while uniting Nagelian reduction and functional reduction in a single model. However, I do see a major problem with it, which so far has gone unnoticed and which to expound will be my focus in the remainder of this paper.

## 2. The Problem

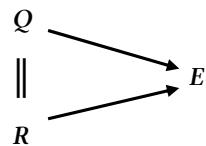
To see the problem and where it appears in SR, recall once more Esfeld et al.’s main goal: They seek to evade multiple realization and simultaneously to deliver a model of reduction which marries the merits of Nagelian reduction with those of functional reduction. On a more general level they also aim, following claim (C), to account with their model both for epistemological *and* ontological reduction.

Bearing this in mind, it appears quite sensible to regard as the key link in the complex chain of reasoning described in the previous section the introduction of functional reduction—since that is the point at which epistemological and ontological models meet most overtly. I believe that this link demands greater attention, and in the following, I will therefore reconsider how exactly functional reduction is supposed to work, what it is originally intended to achieve, and how it reacts when Esfeld et al. combine it with Nagelian reduction.

To start, note again the aim of functional reduction, which is to reduce higher-level properties to lower-level ones. At the outset, this aim is an entirely ontological one, involving no talk of the theories which might seize and describe the properties in question. This is reflected in the fact that the first two steps of functional reduction, *functionalization*, and *realization*, do not refer to theories at all. They rather seem to suggest a purely ontological ‘gaze’ upon the properties under consideration, and those properties’ theoretical representation only begins to matter in step three, where the realizer property is integrated into an explanatory theory.

In light of this, let us recall how exactly functional reduction proceeds. First, a property  $Q$  is functionalized with reference to its characteristic effects  $E$ , yielding a causal role  $H$ . Second, some property  $R$  is found, which realizes  $H$ —whereas  $R$ ’s realizing  $H$  here simply amounts to  $R$ ’s causing  $E$ . As a result,  $Q$  and  $R$  are then reductively identified (see Kim 2005, 101–102). (See Fig. 1; Arrows indicate causation, doubled vertical lines indicate identity)

Figure 1:



Given the ‘ontological gaze’ described above, this way of approaching reduction is fairly unproblematic: We may plainly ‘look out’ into the world, find the characteristic effects of  $Q$ , and notice that those same effects

tion of local reduction, see also Esfeld et al. (2012, 22) for a more detailed statement of its main ideas.

13 For their specific arguments, see Esfeld et al. (2012, 222–223).

are also brought about by  $R$ , which prompts us to identify the two properties.<sup>14</sup>

But this process becomes far more complicated once one recalls the use that Esfeld et al. make of the functional model. Within SR, functional reduction is meant to amend Nagelian theory reduction and to pave the way for functional subtypes. In this context, we can no longer sustain the ‘ontological gaze’—for we must now treat all salient properties as being, in principle, seized by concepts belonging to some mature scientific theory or other.

Specifically, this means that in SR we must acknowledge that  $Q$  and  $R$  do not float freely in metaphysical space, so to speak, but are captured by two (presumably disjunct) theories, call them  $T_2$  and  $T_1$  respectively.<sup>15</sup>

What is more, this shift away from a purely ontological perspective on functional reduction immediately impacts the process of functionalization. For we must now ask the following questions: In which theoretical vocabulary should we express  $Q$ ’s functional definition? Which theory’s concepts do we need to look to if we are to describe  $Q$ ’s characteristic effects? Is it  $T_2$ ,  $T_1$ , or both?

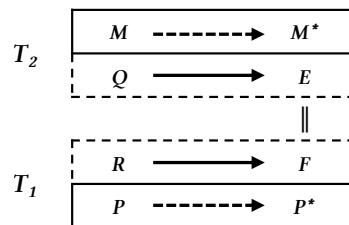
To begin answering these questions, one may note that it is highly sensible to hold that the characteristic effects  $E$  of  $Q$  should be described by the same theory which describes  $Q$  itself. Esfeld et al. also admit this. They claim that “one defines the property types in the domain of [ $T_2$ ] in a functional manner by indicating notably their characteristic effects in terms of [ $T_2$ ]” (Esfeld et al. 2012, 220; cf. Esfeld and Sachse 2007, 3). Additional plausibility may be given to this claim by the following reasoning: If mature theories capture with their laws genuine causal connections, and if properties have unique causal powers which are responsible for the obtaining of these connections (cf. Kim 2005, 18–19; 1976, 162), then those causal connections in which  $Q$  stands with its typical effects—if they genuinely exist—should be captured by the same theory which captures  $Q$ , i.e. by  $T_2$ . Were this not the case, then the concept representing  $Q$  in  $T_2$  would fail to seize one of  $Q$ ’s essential features—namely its unique causal powers. Therefore,  $Q$ ’s functional definition is to be given in terms of  $T_2$ , meaning that  $E$  will be seized by  $T_2$ -concepts.

Is  $E$  also seized by  $T_1$ -concepts? I do not believe so, at least not in the case of genuinely disjunct theories. For if  $E$  were described both by  $T_2$  and  $T_1$ , this would im-

ply that either the *same* concept seizes  $E$  in both theories—in which case the two theories would not be genuinely disjunct;<sup>16</sup> or that *different* concepts seize  $E$  in both theories. But the latter is unsatisfactory insofar as under the assumption that theoretical concepts represent genuinely existing properties, a difference in concepts would usually be thought to imply a difference in properties represented. And awareness of the fact that two concepts originating from different theoretical contexts *actually* seize the same property is exactly the *result* of successful reduction and therefore cannot be one of its presuppositions, lest it make the reductive argument vacuous. If  $E$  is seized by  $T_2$ -concepts it thus cannot, ahead of reduction, also be seized by  $T_1$ -concepts.

Following along the steps of functional reduction, one will quickly see that this adjustment to the process of functionalization also directly affects *realization*: Recall that we are supposed to find the lower-level realizer of the functional role  $H$  by looking for a  $T_1$ -property  $R$  which typically causes  $E$ . But if  $E$  cannot be captured by  $T_1$ -concepts, and if the characteristic effects of  $R$  are described entirely within  $T_1$ ,<sup>17</sup> this would imply that *prima facie*, the characteristic effects of  $Q$  and  $R$  are not identical. Thus, the situation no longer resembles the one pictured in Fig. 1, where  $Q$  and  $R$  both cause  $E$ . Rather,  $Q$  causes  $E$ , and  $R$  causes  $F$ —whereas the former are captured by concepts  $M$  and  $M^*$  in  $T_2$ , and the latter are captured by concepts  $P$  and  $P^*$  in  $T_1$  (see Fig. 2; Broken arrows indicate laws or law-like generalizations):

Figure 2:



As pictured above, this means that now the *identification* of the characteristic effects  $E$  of  $Q$  with the characteristic effects  $F$  of its realizer  $R$  has become a central aspect of realization. For without this identification, no  $T_1$ -property could ever be shown to cause  $E$ —which would mean that the causal role  $H$  could never be realized and that functional reduc-

<sup>14</sup> This is of course somewhat oversimplified.

<sup>15</sup> It is of course possible that  $Q$  and  $R$  happen to be described by the same theory, but such cases are not presently of interest.

<sup>16</sup> Unless we assume the possibility of a theory-independent observational vocabulary.

<sup>17</sup> For reasons similar to those given above for  $E$ ’s being captured exclusively by  $T_2$ .

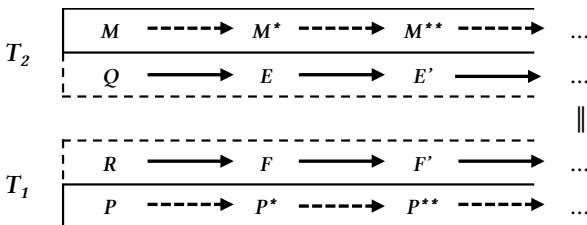
tion, could not succeed. Interestingly, Esfeld et al. reach a very similar conclusion:

“Hence, in order to be in the position to discern in step 2 a certain configuration of properties in the domain of  $[T_1]$  as a realizer of a property type in the domain of  $[T_2]$ , one has to create a link between concepts proper to  $[T_1]$  and concepts proper to  $[T_2]$ . This link is generally established on the basis of *identifying the effects* brought about by a certain property in the domain of  $[T_2]$  in a given situation with effects brought about by a configuration of a certain type in the domain of  $[T_1]$  in that situation.” (Esfeld et al. 2012, 220. Emphasis added)

However, Esfeld et al. do not continue this train of thought any further. After the cited passage, they move on to an example of functional reduction and do not return to the matter of identifying  $Q$ 's and  $R$ 's characteristic effects.

I believe that this is a grave oversight. Because the identification of  $E$  and  $F$  is not a trivial matter, merely to be mentioned once and then taken for granted. For in the present context, the only way of establishing the required (reductive) property-identity between  $E$  and  $F$  is exactly *through functional reduction*. Recognizing this is of the utmost importance, as it implies that the functional reduction of  $Q$  (i.e.  $Q$ 's reductive identification with its realizer  $R$ ) presupposes the previous functional reduction of its characteristic effects  $E$  (i.e.  $E$ 's identification with its realizer  $F$ ). But if functional reduction indeed proceeds as here illustrated, this means that the functional reduction of  $E$  in turn presupposes the functional reduction of the characteristic effects  $E'$  of  $E$ , and so on. We are thus off on a ‘sideways’ regress, where the functional reduction of each property depends on the functional reduction of those properties which lie ‘ahead’ of it in the causal chain (see Fig. 3).

Figure 3:



It remains to be determined whether this regress is vicious. I believe that it is. To see why, consider the properties to be reduced. Either the genuine causal chains (or networks) which connect these properties (and/or property tokens) in the actual world are infinite, in which case functional reduction can never succeed as there are always more effects to be functionalized; or the chains are finite, in which case functional reduction cannot succeed either, since the last ‘link’ in these chains will always turn out to be an epiphenomenon and thus not admit of functionalization.<sup>18</sup>

Thus, there is a fundamental problem in the form of a vicious regress which emerges in SR once functional reduction is combined with Nagelian theory reduction. Furthermore, this problem is detrimental to SR because the functional subtypes Esfeld et al. envision are built on the prior possibility of identifying all possible realizers of a certain functionalized property. If, for the above reasons, it is no longer possible to identify *any* realizers of the targeted property, one will not be able to build functional subtypes and SR will fail.

### 3. Possible Objections and Reductive Explanation

Having stated what I perceive to be SR's central problem, I would now like to anticipate two objections. The second of these will demand greater attention and will also lead into a further aspect of the problem at hand. As a first objection, one could argue that at least one of the assumptions which lead to the abovementioned regress were mistaken and that hence, the regress does not actually arise. For this purpose, one could for example, despite the arguments given above, deny that the characteristic effects  $E$  of  $Q$  are captured exclusively by  $T_2$ -concepts and argue that they are captured exclusively by  $T_1$ -concepts instead. Were this correct, realization could proceed as originally intended, since if  $Q$ 's characteristic effects were described within  $T_1$ , it would in principle be possible to find a  $T_1$ -property which has these same effects.

I believe that the main force of this objection can be mitigated by pointing to the fact that if it were true that  $Q$ 's characteristic effects are described by  $T_1$ , then we would be dealing with an instance of what

<sup>18</sup> Even if we were to include in the functional definition of a property its typical *causes*, we would not evade the problem. For assume, e.g., that the functional reduction of the epiphenomenon  $G$  presupposes the functional reduction of its typical cause  $K$ . In this case, the functional reduction of  $K$  in turn presupposes on the one hand the functional reduction of its characteristic effect  $G$ —which leads into circularity; and on the other the functional reduction of its characteristic cause  $L$ —which leads again into the regress at hand.

Jaegwon Kim calls “downwards causation”, where upper-level properties cause lower-level ones (Kim 2005, 44.). Downward causation is, however, ruled out quite forcefully by the causal exclusion argument which, recall, Esfeld et al. do accept (see Esfeld and Sachse 2007, 1-2). E’s being seized only by  $T_I$ -concepts is therefore out of the question. A possible rejoinder to this is to claim that upon successful functional reduction, downwards causation is identical with lower level causation and thus does not pose a problem. But again, this is the case only *after* reduction has succeeded. It is a result of functional reduction that downward causation (and also upper-level causation) is identical with lower-level causation—but it cannot be presupposed as a basis for *carrying out* functional reductions.

As a second objection, one could argue that while the vicious regress which constitutes the problem at hand does indeed arise if the only way of establishing an identity between  $E$  and  $F$  is through functional reduction, there are other ways of arriving at this identity. One could make recourse, for example, to non-causal identity criteria for properties. Or, more pragmatically, one could claim that it is enough to simply *postulate* such reductive identities, which are then vindicated when they lead to useful and empirically traceable reductions. I have some sympathy with this objection. Indeed, when I stated above that SR may be a workable model, I was speaking exactly of this approach to evading the problem and rescuing SR. But I do nevertheless have my reservations, especially concerning the latter ‘pragmatic’ approach, which I will detail in the closing passages of this section. To start, consider what the proponents of functional reduction take to be its main benefits. Jaegwon Kim, for instance, insists that functional reduction is preferable to Nagelian reduction or to the mere postulation of reductive identities on the grounds that, unlike the latter two, it can provide *reductive explanations* (see Kim 2005, 99-102 and 111-112). In the context of the mind-body problem, in which Kim situates his model, this means that functional reduction cannot only, for example, state that pain is identical to C-fibres firing but that it can also adequately explain *why* this reductive identity holds. Why is pain identical with C-fibres firing? Because C-fibre activation and pain have the same characteristic effects. Functional reduction—as envisioned by Kim—is therefore able to bridge the ‘explanatory gap’ that often plagues the reduction of mental properties.<sup>19</sup>

---

<sup>19</sup> It is important to note, as Kim himself does, that functional reduction cannot bridge this gap *completely* (see Kim 2005, 170 and 173-174).

But Kim’s model can provide reductive explanations only because functional reductions do not invoke any other *unexplained* property-identities or correlations in their premises. Indeed, highlighting this very same point, Kim’s critique of Nagelian reduction amounts exactly to the observation that the bridge-laws required by the Nagelian model do not, in fact, have any explanatory power because they themselves remain unexplained:

“What was being demanded was an explanation of why pain correlates with Cfs [C-fibre stimulation] to begin with. Why is it that pain arises when Cfs occurs, not when A $\delta$ -fibres are activated? [...] These are precisely requests for explanations of the bridge laws themselves. In using the bridge laws as auxiliary premises of reductive derivations, the Nagelian reductivist is simply assuming exactly what needs to be derived [...].” (Kim 2005, 100)

Yet in the context of SR, successful functional reductions must invoke *precisely* such unexplained identities: If the functional reduction of  $Q$  to  $R$  can only succeed by postulating as a premise the identity of  $Q$ ’s and  $R$ ’s characteristic effects, then the burden of explanation is simply shifted from the identification of  $Q$  and  $R$  to that of  $E$  and  $F$ . On SR’s approach to functional reduction, the answer to the question: “Why is pain identical to C-fibres firing?” can therefore only be: “Because we *postulate* that C-fibre activation and pain have the same characteristic effects”. And this certainly does not bridge the explanatory gap.

So, even if we assume that the second objection to the vicious regress succeeds and that the identity of  $E$  and  $F$  may simply be postulated, SR sustains a serious blow (though it remains a workable model).<sup>20</sup>

## Concluding Remarks

Summarizing the preceding arguments, we may notice that we have ended up facing the following dilemma: Either we acknowledge the problem I have tried to illustrate and accept that SR fails because of the resulting vicious regress; or we deny that the regress arises and claim that we may rescue SR by postulating (or otherwise arriving at) the needed property identities—in which case SR remains workable but loses most of its explanatory power.

---

<sup>20</sup> I have here left out of the discussion the possibility of there being other strategies of identifying properties while at the same time providing reductive explanations. Should there be any such strategies, the situation would need to be reconsidered.

Both horns of this dilemma are unsatisfactory, though supporters of SR will of course find the second horn far more palatable. In fact, even if one embraces the second horn and accepts that SR can no longer provide reductive explanations, it may still be preferred to other models of reduction since as long as it remains workable, it can indeed—after all—escape multiple realization.

However, as anticipated in the introduction, the above dilemma does not only affect SR. Our central problem arises not from functional subtypes, but rather from the previous combination of Nagelian theory reduction and functional reduction. This means that all present or future models which attempt to likewise combine these two models will face the very same dilemma. Thus, the problem presented here is a problem for *all* functional models of theory reduction.

Moreover, since on a more general level the problem arises directly from Esfeld et al.'s claim (C) which states that ontological and epistemological reduction entail one another, the problem even affects functional reduction directly. For if all models of ontological reduction must, following (C), be reinterpreted as also covering epistemological reduction, so must functional reduction—and then the problem arises on its own, even if the Nagelian model is left out of the picture.

Finally, this also suggests that claim (C) needs to be given more extensive consideration. After all, we may point out that the problems which plague the various models of reduction here discussed all result more or less directly from (C). Multiple realization troubles Nagelian reduction only when one considers the ontological implications of bridge-laws; the elimination of higher-level theories is only implied by local functional reduction once it is framed in the context of theory reductions; and likewise, the vicious regress here shown to threaten SR and functional reduction. Of course, this does not mean that the present paper should retroactively be read as a *reductio* of claim (C). But it should open up for future discussion the relationship between epistemological and ontological models of reduction.

## References

- Block, Ned. 1980. "Introduction: What is Functionalism." In *Readings in Philosophy of Psychology, Volume I*, edited by Ned Block, 171-184. Cambridge, MA: Harvard University Press.
- Bregant, Janez. 2006. "The Limits of Functional Reduction." *Synthese Philosophica* 43, 219-229.
- Eronen, Markus I. 2011. "Replacing Functional Reduction with Mechanistic Explanation." *Philosophia Naturalis* 48, 125-153.
- Esfeld, Michael and Christian Sachse. 2007. "Theory Reduction by Means of Functional Sub-Types." *International Studies in the Philosophy of Science* 21, 1-17.
- Esfeld, Michael, Christian Sachse and Patrice Soom. 2012. "Marrying the Merits of Nagelian Reduction and Functional Reduction." *Acta Anal* 27, 217-230.
- Fodor, Jerry A. 1974. "Special Sciences (Or: The Disunity of Science as a Working Hypothesis)." *Synthese* 28, 97-115.
- Heil, John. 2016. "Levels of Reality." *Ratio* 16, 205-221.
- Kim, Jaegwon. 1976. "Events as Property Exemplifications". In Myles Brand and Douglas Walton *Action Theory*, edited by M. Brand and D. Walton, 159-177.
- . 1998. *Mind in a Physical World*. Cambridge, MA: MIT Press.
- . 2005. *Physicalism, or Something Near Enough*. Princeton, NJ: Princeton University Press.
- Kistler, Max. 2005. "Is Functional Reduction Logical Reduction?" *Croatian Journal of Philosophy* 5, 219-234.
- Lewis, David. 1966. "An Argument for the Identity Theory." *Journal of Philosophy* 63, 17-25.
- . 1972. "Psychophysical and Theoretical Identifications." *Australian Journal of Philosophy* 50, 249-258.
- . 1981. "Mad Pain and Martian Pain". In *Readings in the Philosophy of Psychology*, edited by Ned Block, 216-222. Cambridge, MA: Harvard University Press.
- Marras, Ausonio. 2000. "Critical Notice." *Canadian Journal of Philosophy* 30, 137-160.
- . 2005. "Consciousness and Reduction." *British Journal for the Philosophy of Science* 56, 335-36.
- Nagel, Ernest. 1961. *The Structure of Science. Problems in the Logic of Scientific Explanation*. New York and Burlingame: Harcourt, Brace & World.
- Sarkar, Sahotra. 2015. "Nagel on Reduction." *Studies in History and Philosophy of Science* 53, 43-5

---

**Manuel Steffen (23)** schliesst im kommenden Semester sein Bachelorstudium ab. Er interessiert sich für verschiedenste Fragen der theoretischen Philosophie, beschäftigt sich momentan aber vor allem mit Modalität, Kausalität und Metaphysik im Allgemeinen.

# Sex Work is Work

Arguments against viewing sex work as a  
morally distinctive labour market

## Introduction

In her 2010 book *Why Some Things Should Not Be for Sale: The Moral Limits of Markets*, Debra Satz establishes a framework to determine market failure which exceeds both classical economic models which usually determine market failure according to a market's efficiency, as well as a liberal egalitarian theory which deals with market failure through (re)distribution (Satz 2010, 5). By contrast, Satz follows a "nuanced view of the idea of market failure, one that takes into account how markets shape our relationships with others in ways that goes beyond the idea of unabsorbed economic costs" (Satz 2010, 93). In this context, Satz argues "that some markets are *noxious* and need to be blocked or severely constrained" (Satz 2010, 93, emphasis mine) in order to avoid the negative outcomes for the people involved or for society at large. The scope of this paper lies in a critical assessment of the specific case of a market in sexual services that Satz analyses when she looks at specific markets and their moral ramifications. In line with Satz's previous theorisation on the moral limits of markets, Satz concludes that a market in sex work<sup>1</sup> is morally wrong because "the practice of contemporary prostitution contributes to and also embodies the perception of women as socially inferior to men" (Satz 2010, 135). Although I generally agree with Satz's assessment of the consequences in terms of policy related to sex work, I argue that Satz's moral assessment of a market in sex work has to be questioned. Specifically, I argue that Satz's arguments for the distinctiveness of sex work compared to other labour markets do not withstand

scrutiny. Furthermore, I argue that Satz underestimates the power of the individual choice to engage in sex work. Rather than viewing sex workers as the embodied reproducers of a culture that oppresses and objectifies women, I will suggest that we should pay attention to how sex workers perceive themselves in what they do and what that means for a sexist and heteropatriarchal culture of objectification which Satz criticises.

In order to achieve the task outlined above, I will first summarise Satz's argument for the noxiousness of a market in sex work. In the following section, I will try to show that Satz's arguments for the distinctiveness of sex work compared to other labour markets stand on relatively weak empirical ground and are themselves morally problematic. By arguing that sex work is not a distinctive labour market when it comes to its propensity to portray a picture of women as socially inferior to men, I hope to provide a differentiated approach which takes into account the labourer's perspective and does justice to their agency.

## Satz's Argument for the Noxiousness of Sex Work

This section will first provide a brief overview of Satz's general account on the moral assessment of markets. Secondly and more extensively, it will provide a summary of Satz's evaluation of a market in sex work.

Debra Satz provides a refreshingly differentiated account on the moral limits of different markets. What makes her account especially valuable is its straightforward applicability and the fact that it nonetheless remains remarkably attentive to detail. The attentiveness to detail is what particularly sets Satz's account apart from (most) economic accounts on the limits of markets, especially because she views labour markets differently from other

<sup>1</sup> Satz repeatedly uses the term "prostitution" or "prostitute" to refer to sex workers. Wherever it is possible, I will avoid these terms since they refer to the description of an identity related to the policing of sexual deviance which originates in the nineteenth century, roughly the same period and context that the term "homosexual" came about (Grant 2014, 13ff.). By contrast, the term "sex worker" and "sex work" was coined by sex workers themselves and relates to the service they provide, rather than to their identity (*ibid.*).

markets (e.g.: markets in goods).<sup>2</sup> Generally, Satz understands *markets* as institutions where repeated transactions between equally entitled parties take place under the condition of *freedom of choice* (to agree to or refuse a transaction) and *competitiveness* (“make another offer or strike a better deal with someone else” (Satz 2010, 15)). The repetitiveness of transactions is especially important as a matter of definition, since, in Satz’s understanding, single market transactions “can be noxious without there being a noxious market” (Satz 2010, 15). Analytically, *markets* have to be understood separately from the *market system* which is only “the abstraction that is supposed to link the set of all such [specific] markets” (Satz 2010, 16). Satz elaborates that markets are always socially embedded and cannot be viewed separately from the values and the context within which they exist. Hence, markets represent a matter of “*public* concern of all citizens, whether or not they directly participate in them” (Satz 2010, 16).

Satz argues that “[s]ome markets produce *extremely harmful outcomes* not only for individuals but also *for society*” because “they undermine the framework needed for a society of equals” and/or support “relations of humiliating subordination or unaccountable power” (Satz 2010, 9). Satz defines markets which produce these outcomes as *noxious markets*. There are, more precisely, four criteria to determine the noxiousness of a particular market, two of which refer to the consequences of these markets and two refer to their sources. Each of the two groups of criteria is subdivided into individual and collective extreme harms. For present purposes, only the criterion of collective consequences will be presented.<sup>3</sup> This is because Satz uses only this criterion to argue for the noxiousness of a market in sex work. It concerns the extremely negative consequences a particular market can have for society as a whole. Satz states that “[t]he operation of these markets can undermine the social framework needed for people to interact as equals, as individuals with equal standing” (Satz 2010, 95). To “interact as equals” has to be understood in a minimal sense, i.e. having “the preconditions necessary for individuals to make claims on one another and interact without having to beg or to push others around” (Satz 2010, 95). For present purposes it has to suffice to say that the ramifications of this condition go as far as opening the door to undermining the values of democratic societies as such because

interacting as equals is not only one of the important preconditions of markets (as Satz defines them), but also of democracies. That is why “the democratic state has an interest in withholding its support from institutions that cultivate subordination and servitude, even if those institutions are not strictly illegal” (Satz 2010, 104).

For Satz, a market in sex work is such a case. Satz argues that the moral problem of a market in sex work lies therein that “the practice of contemporary prostitution contributes to and also embodies the perception of women as socially inferior to men” (Satz 2010, 135). In other words, a market in sex work embodies and reproduces the perception of the “class of women” (Satz 2010, 144) in contemporary societies as people with inferior social standing and produces an extremely harmful outcome for society, thus fulfilling Satz’s collective consequence criterion for a noxious market. According to Satz, the damage that is done to all women by a market in sex work lies in “conceiving of a class of women as needed to satisfy male sexual desire” and thus representing “women as sexual servants to men” (Satz 2010, 144). Satz makes this argument after she dismisses a number of essentialist arguments against sex work, one of which merits mentioning because Satz grounds her own argument on it.

The essentialist argument, as brought forward by Elizabeth Anderson, states that sex workers cede their bodies in the exchange and that this constitutes a harm to “the nature of sex as a shared good, based on the recognition of mutual attraction” (Satz 2010, 142). Within this exchange, both parties of the exchange lose the entitlement of having sex as a shared good, but, according to Anderson, it is worse for the sex worker because she<sup>4</sup> cedes her body and the customer only cedes cash (Satz 2010, 143). But how is that different from any other kind of job in the labour markets? As Satz rightly stresses, it remains unclear why this sort of exchange would constitute anything more harmful than many other widely accepted exchanges (Satz 2010, 143). This seems to be a very important objection. It has to be stressed that there is an abundance of jobs that are only taken up due to some degree of desperation and the need for money. Furthermore, the renting of parts of our bodies (and minds)

<sup>2</sup> Satz adopts this argument for the distinctiveness of labour markets from other markets from Karl Polanyi’s central conclusions in his monumental work *The Great Transformation* (2001 [1944]).

<sup>3</sup> The whole elaboration of Satz’s criteria can be found in Satz 2010, 94-97.

<sup>4</sup> It is important to note that Anderson and Satz always assume that sex workers—or at least the statistically and morally significant proportion of sex workers—are cis-gendered women. This is why I use the she/her pronouns when referring to the arguments of Satz and Anderson. However, I have to stress that this portrayal of the demo-graphic within the field of sex work is highly inaccurate and misleading, especially when we consider the dis-proportionate amount of trans and gender non-conforming sex workers. However, a conclusive argument against Satz’s assumptions about sex workers’ sex, gender and sexuality would exceed the scope of this paper.

is the most common form of how labour markets work. Indeed, I cannot think of any particular labour market that does not involve that kind of transaction. The sex worker and activist Carol Queen sums this up very well when she says that “we [sex workers] are not *selling* ourselves or our bodies [...] any more than does any worker under capitalism”, and that “any worker under capitalism is subject to mistreatment” (Queen 1997, 135). Accordingly, it appears like the issue of mistreatment and the renting of parts of our bodies is rather something that stems from the abstract market system that Satz distinguishes from particular markets. It might be another issue if people were actually losing some bodily capacity in the exchange, as is the case with markets in organs, body parts and the like. However, this is not the case with sex work. According to proponents of this kind of essentialist argument, sex workers lose their dignity because of the desperation of their exchange. This is not only ignoring the desperation of the need to work for most people in any capitalist society, it is also overtly denying any agency to sex workers. We will return to that element of agency denial in the next section. For now, I will close this section with the missing details of Satz’s modified account on the moral noxiousness of a market in sex work.

For her argument for the noxiousness of sex work as a labour market, Satz modifies the *degradation objection*, as she calls the latter essentialist argument by Anderson. Instead of focusing on an alleged weak agency and on the alleged degradation of the sex worker, Satz draws “a connection between the current practice of prostitution and the lesser social status of women” (Satz 2010, 143). Satz does not assume that this effect is something inherent in sex work but is rather a product of sex work within a culture of gender inequality, in which women are oppressed and objectified in myriad ways (Satz 2010, 143). Satz distinguishes two separate types of gender inequality which must not be conflated, even if they “clearly interact” (Satz 2010, 146). Firstly, there is distributional inequality. It involves unequal income, unequal rates of poverty<sup>5</sup> and unequal division of unpaid labour in the household (Satz 2010, 144f). For Satz, this first dimension is not directly affected by the market in sex work (Satz 2010, 146). Yet, sex work might be the only work field where women typically earn more than their male colleagues.<sup>6</sup> Secondly, there is gender inequality related

to the standing of women in society. Satz calls this dimension “status inequality” (Satz 2010, 146). This is the dimension of gender inequality that Satz sees exacerbated by the practice of contemporary sex work because “prostitution is a theater of inequality; it displays for us a practice in which women are seen as servants of men’s desires” (Satz 2010, 147) and this has a negative effect “on how men perceive women and on how women perceive themselves” (Satz 2010, 146). If we accept these claims, the question arises how this negative effect on the standing of women is any different in sex work compared to other professions which are typically done by women or in which the gendered division of jobs would equally have a different outcome for women and men, like secretarial labour, fashion modelling, or housecleaning. Satz provides three reasons why sex work is a distinctive labour market when it comes to its propensity to embody and reproduce an image of women as a class of inferior social standing.

In the following section, these three arguments for the distinctiveness of sex work, henceforth called *distinctiveness arguments*, will be presented and discussed one by one. It will be argued that Satz’s arguments do not show a distinctiveness of sex work when it comes to the alleged extreme harm for society. Satz does not clarify if the three distinctiveness arguments are necessary and sufficient or if they are individually sufficient. However, I will argue against all three arguments and will thus try to show that there is in fact no reason to view sex work distinctively to other forms of labour when it comes to a potential harm to society or the class of women. Instead, I will argue that Satz’s arguments have an underlying misinterpretation of the social reality they analyse and that this results in an underestimation of the individual power to challenge the heteropatriarchal culture that Satz sees exacerbated by sex work in the first place. Rather than saying that sex work is morally noxious because it exacerbates gender status inequality for all women, I argue that there is even a potential that originates within this work in terms of changing the image of women as passive sexual objects.

## Critical Assessment

As a first argument for the distinctiveness of sex work, Satz invokes consequences of a general moral view about sex workers: “First, a large number of people currently believe that prostitution, unlike housecleaning, is especially objectionable. Holding such moral views of prostitution constant, if prostitution continues to be primarily a female occupation, then the existence of prostitution will disproportionately fuel negative images of women”

<sup>5</sup> As Satz rightly stresses, unequal rates of poverty are even worse if we look at intersectional discriminations, like the position of women of colour and, I would add, transgender PoC (Satz 2010, 144).

<sup>6</sup> This information is taken from the sex worker Julian Marlowe and reflects his personal experience in the field of sex work (Marlowe 1997, 143).

(Satz 2010, 149). I argue that this does not ground a distinctiveness of sex work compared to other labour market interactions because it misinterprets the social reality in which the image or stigma operates. It might indeed be true that most people think sex work is especially objectionable and that as long as women continue to do most of sex work, all women are affected by the stigma. However, the stigma of the sex worker as the promiscuous object for male use, henceforth called whore *stigma*<sup>7</sup>, is not limited to sex workers. It is applied to any woman<sup>8</sup> who appears to live a promiscuous sexual life or who is not conforming with established norms about sexuality, gender and family. Consider what the sex worker Carol Queen has to say about the stigma that Satz sees reproduced in the contemporary practice of sex work. She says that women “are supposed to exist sexually within a (married, monogamous) relationship with a man, or else not at all” (Queen 1997, 127). It is not by virtue of the existence of sex work that the whore stigma is applied to women but by virtue of “compulsory heterosexuality and ‘good girl’ propriety” (Queen 1997, 127). This becomes especially apparent when we consider the policing and punishment of presumed sex workers. The sex worker and activist Jill Nagle fittingly says that “one does not have to actually *be* a whore to suffer a whore’s punishment or stigma. [...] Recent laws [...] give police officers the power to arrest someone for *appearing to intend* to exchange sex for money” (Nagle 1997, 5). This practice has profound effects on already marginalised groups of people, like transgender women.<sup>9</sup> I do not think that Satz would agree that we should hold morally accountable any woman who *appears to intend* to exchange sex for money or a woman who is perceived to have a promiscuous lifestyle and is thus being called a whore or a slut for reproducing an image of all women as objects for male sexual pleasure. Equally, we would not want to hold morally accountable lesbian women and queer people for being blamed for not conforming to compulsory heterosexuality. To do that

is to address the wrong side of the problem.

In response, Satz might object that if there were no sex workers, the whore stigma would not exist and would therefore also not be applied to other women who do not conform with compulsory heterosexuality and good girl propriety. I argue that it is far more likely that, under the assumption of a non-existence of sex work, stigmatisation would remain because the sexual and gendered norms would still exist and deviance from these norms would still be socially sanctioned in the form of some sort of whore stigma.

To sum up, the first argument for the distinctiveness of sex work in its role of embodying and reproducing an image of women as the sexual servants of men does not plausibly clarify the alleged direction of causality in its interpretation of the social reality within which the stigma operates. Instead of making sex work responsible for the production of an image of women as sexual servants of men and thus for reproducing the dimension of gender inequality that concerns the standing of women in society, it can be plausibly argued that the image of women as sexual servants to men is a product of a culture of gender inequality that sets clear distinctions what is sexually acceptable and what is not.

The second argument for the distinctiveness of sex work also does not withstand scrutiny. Satz states that “prostitution represents women as objects for male use. [...] [P]rostitutes are far more likely to be victims of violence than other professions; they are also far more likely to be raped than other women” (Satz 2010, 149).

I agree with both of the above statements. However, it does not follow from them that sex work is somehow exclusively distinctive to other labour markets. Firstly, sex work is not so different from other labour markets like fashion modelling, the advertisement or the movie industry with respect to its propensity to project an image of women as objects for male use. Secondly, it is problematic to ground a distinctiveness of sex work on the empirical observation that sexualised violence and abuse is more likely to occur in the field of sex work. Just because the statistical probability of being abused in the labour market of sex work is significantly higher, it does not necessarily follow that sex work is essentially distinctive to other labour markets. This argument essentialises the experience of the sex worker in that it suggests that there is something fundamentally different to being a sex worker than to being a female worker in other fields. This essentialising feature is especially apparent when Satz states that “a prostitute’s “no” does not, to the male she services as well as to other men, mean no” (Satz 2010, 149). The implication of the latter statement is that,

<sup>7</sup> This term is taken from sex worker activists like Priscilla Alexander (1997), Carol Queen (1997), or Jill Nagle (1997).

<sup>8</sup> To speak of ‘women’ in this context is actually quite inaccurate and misleading. Rather, we should speak of sexually marginalised people because the stigma is not strictly limited to women, just as much as the policing is targeted at “people whose sexuality and gender are considered suspect” (Grant 2014, 9; cf. footnote 40). For the treatment of gender just language in this paper consider footnote 19.

<sup>9</sup> The sex worker and activist Melissa Gira Grant writes: “People who are profiled by cops as sex workers include, in disproportionate numbers, trans women, women of color, and queer and gender non-conforming youth. This isn’t about policing sex. It’s about profiling and policing people whose sexuality and gender are considered suspect” (Grant 2014, 9).

because there is a higher prevalence of abuse and rape in the field of sex work, sex work is distinctive from other labour markets because, unlike other workers, sex workers lose their agency to say “no” by virtue of being sex workers. This not only ignores that some sex workers are in a position to choose their clients and determine the rules and boundaries of their transaction. It ignores how common sexual abuse and rape is in many labour markets. We have known this long before the #MeToo movement but we are getting a clearer idea of how widespread and common sexual abuse in the workplace and everywhere is ever since. In fact, the commonness of rape in our modern societies goes as far as having to be called *rape culture*<sup>10</sup>: a culture “in which rape and sexual humiliation of women and girls is so normalised that it does not register as a crime in the minds of the assailants,”<sup>11</sup> as Laurie Penny (2013) says it to the point. To name some of the markets and institutions that were revealed to be especially prone to this kind of abuse during the 2017 and ongoing #MeToo movement, consider the film, fashion, media, music, the financial and sports industry, politics and academia all over the world. In light of the countless testimonies<sup>12</sup> of abused women in this vast diversity of labour markets, it is problematic to state that it grounds a distinctiveness of sex work compared to other markets. What does follow from the above observations is that the “no” of many women in many labour markets does not mean “no” to many males in positions of power. Accordingly, the problem lies with the culture of rape and the society that educates men in a way that leads them to think they are allowed to objectify and abuse women with impunity. Sex work is certainly prone to patriarchal abuse and rape is certainly an experience that many sex workers cannot evade. What I try to highlight is that rape, sexual abuse and harassment is the product of power relations which extend way beyond sex work as a labour market.

Finally, in the last distinctiveness argument Satz states that there is a “third-party harm: the effects that prostitution may have on other women’s sexual autonomy. [...] [I]f prostitution was viewed as just another job analogous to other forms of employment, then sex could be included as part of any number of jobs. Women who did

not wish to have sex on demand might find that their employment options were limited and that they were less employable on the labor market” (Satz 2010, 149). This argument for the distinctiveness of sex work does not withstand scrutiny either. In light of what has been said in the last paragraph, saying that sex work has to be viewed distinctively from other labour markets because sex work’s normalisation as just another job would affect the class of women as a whole because “then sex could be included as part of any number of jobs” and that would limit women’s job options, appears to be grounded on a romanticised notion of what the reality for women at least in some other labour markets currently is. This is not to say that an expectation of sex is a part of any job. However, in order to contradict Satz’s argument, there only has to be sufficient evidence that there already is such an expectation in other labour markets. This obviously being the case<sup>13</sup>, it is hard to argue that such an expectation would stem from viewing sex work as just another job because, currently, sex work is not viewed as just another job.

Secondly, to say that the labour market of sex work harms all women because it generates a general expectation for sex in any labour market is ignoring sex workers’ agency in determining the limits and boundaries of their performances as sexualised objects of desire. If one argues that sex work projects an image of women as sexual servants to men, one has to assume that sex workers do not actually choose who they serve and under which conditions. In other words, there is an assumption of weak agency on the sex worker’s side. This assumption is evident when we again consider Satz’s statement that “a prostitute’s “no” does not, to the male she services as well as to other men, mean no” (Satz 2010, 149). This (general) statement contradicts Satz’s earlier statement that sex work does not “necessarily involve weak agency” (Satz 2010, 136). More importantly, it is a highly problematic assumption by itself because false assumptions about weak or nonexistent agency are a common feature of the heteropatriarchal and sexist culture that Satz sees exacerbated in the practice of contemporary sex work. Conversely, one could argue that viewing sex work as a job like any other might even have an opposite effect than what Satz claims. This is due to the fact that, ideally, sex workers decide to sell the performance of sex and clearly set the conditions and boundaries for any transaction with their customers, thereby cultivating a norm

---

10 Further reading on and evidence of rape culture: Fanganel 2019.

11 I deliberately chose an article from before the #MeToo movement to underpin the fact that we have known about rape culture for much longer and #MeToo only made clear exactly how common and widespread rape and sexual abuse is in virtually every labour market one can think of.

12 To provide just one of countless numbers on the movement: within 24 hours, 4.7 million people from all around the world used #MeToo on Facebook alone (Pengelly 2017).

---

13 Martin Pengelly from The Guardian wrote: “Within days [of the launch of the #metoo movement], millions of women and some men used Twitter, Facebook and Instagram to disclose the harassment and abuse they had faced in their own lives” (Pengelly 2017).

of consent. Obviously, there are single transactions that do in fact involve weak agency and are driven by a general desperateness of the need to work. These transactions could indeed project an image of women as needing to serve male sexual desire. But according to Satz, single market transactions “can be noxious without there being a noxious market” (Satz 2010, 15). Accordingly, one cannot infer a general lack of agency from some instances where this actually occurs. If one conversely adopts a narrative which does acknowledge sex worker’s capacity for informed consent, the image that is embodied and reproduced through the contemporary practice of sex work is not one of portraying women as sexual servants of men but one of cultivating the need to ask and get consent for any given intimate interaction between human beings. We would be at an extremely different point with rape culture if all heterosexual cis-males were taught to get consent before they touch someone or engage in sexual activities. However, this counter-narrative can only be adopted if we acknowledge sex workers’ capacity to say “no” to a transaction.

To sum up, none of the distinctiveness arguments are convincing. It is most certainly objectionable when women are portrayed to be the sexual servants of men and I agree that this has ramifications which go as far as putting in question the claim of living democratic values. However, this latter process which Satz describes is not distinctive of sex work but is rather the product of a more complex and pervasive system of social control, i.e. the cis-hetero patriarchy.

## Conclusion

The purpose of this analysis has been to critically examine Debra Satz’s argument for the noxiousness of a labour market in sex work. After having established that Satz’s argument critically relies on the claim that there is something distinctive to a market in sex work, it has been argued that none of the distinctiveness arguments withstand scrutiny. Accordingly, it has been argued that the imagery of women as sexualised objects for male use is not something distinctive to the labour market in sex work and that, conversely, the underlying culture of male dominance could even be mediated if we would accept a market in sexual services as a labour market like any other, due to the norm of consent that ideally underlies sexual market transactions. By exclusively resorting to voices who come from the field of sex work itself, I have tried to underpin the credibility of my claims, since no one knows better how the market works and how public representation affects it than the labourers themselves. By doing that, I have tried to avoid “the production of

feminist discourse around prostitution by non-prostitutes [which] alienates the laborer herself from the process of her own representation” (Nagle 1997, 2). However, much of what has been argued needs further scrutiny. It has not been possible to shed light on the complex mechanisms that embody, produce and reproduce the imagery of women as sexualised objects for male use within the highly problematic culture that Satz rightly criticises and sees as a problem for claims to democratic values like equality. Also, much needs to be said about the proper regulation of a labour market in sex work. Specifically, it should be worked out which regulations would support sex workers’ agency and self-determination. For now, it has to suffice to say that a market in sex work cannot be called a distinctively noxious market, at least according to Satz’s arguing.

## References

- Alexander, Priscilla. 1999. “Feminism, Sex Workers, and Human Rights.” In *Whores and other feminists*, edited by Jill Nagle. New York: Routledge, 83-97.
- Fanganel, Alexandra. 2019. *Disrupting Rape Culture. Public Space, Sexuality and Revolt*. Bristol: Bristol University Press.
- Grant, Melissa Gira. 2014. *Playing the Whore. The Work of Sex Work*. London: Verso.
- Marlowe, Julian. 1997. “It’s Different for Boys.” In *Whores and other feminists*, edited by Jill Nagle. New York: Routledge, 141-144.
- Nagle, Jill. 1997. *Whores and other feminists*. New York: Routledge, 1-18.
- Pengelly, Martin. 2017. “#MeToo: thousands march in LA as sexual misconduct allegations continue.” *The Guardian*, November 12, 2017. (<https://www.theguardian.com/world/2017/nov/12/metoo-march-hollywood-sexual-assault-harassment> [1.7.2019]).
- Penny, Laurie. 2013. “Laurie Penny on Steubenville: This is Rape Culture’s Abu Ghraib Moment.” *NewStatesman*, March 19, 2013. <https://www.newstatesman.com/laurie-penny/2013/03/steubenville-rape-cultures-abu-ghraib-moment> [3.7.2019].
- Polanyi, Karl. 2001 [1944]. *The Great Transformation. The Political and Economic Origins of Our Time*. Boston: Beacon Press.
- Queen, Carol. 1997. “Sex Radical Politics, Sex Positive Feminist Thought, and Whore Stigma.” In *Whores and other feminists*, edited by Jill Nagle. New York: Routledge, 125-135.
- Satz, Debra. 2010. *Why Some Things Should Not Be for Sale. The Moral Limits of Markets*. New York: Oxford University Press.

---

**Mo Brunold** (Pronomen Mo/Mos) studiert politische Philosophie im Master und hat soeben Mos Masterarbeit im Rahmen einer Dekolonialisierung der politisch philosophischen Theoriebildung abgegeben. Mo interessiert sich für radikale politische Theorie, Dekolonialisierung der Wissenschaften und radikalen Queer-Feminismus.

# The Myth of Reverse Discrimination

Does Affirmative Action Discriminate Against the Majority?

## Introduction

With the current rise of alt-right and men's rights movements the view that measures to increase diversity discriminate against the majority, is spreading. Jordan Peterson for instance claims that today, men are the victims of gendered oppression (Lefkovitz 2018). The last years have seen several lawsuits against institutions committed to affirmative action programs claiming "reverse discrimination". Thus, legal journals of Harvard University and New York University were sued "over allegations that they illegally give preference to women and racial minorities" (Binkley 2018). The view that affirmative action is no longer needed and additionally, that it unfairly punishes the majority (for instance quotas for women<sup>1</sup> are discrimination against men<sup>\*</sup>), seems to be spreading. This, however, is not a new phenomenon: Since the 1970s, affirmative action programs have incited heated debates centred around the question whether they discriminate against the majority.

In our everyday language use "discrimination" is very negatively charged. On the face of it, discrimination is an act that distinguishes between persons on the basis of some trait; this type of making a distinction is discrimination in a purely *descriptive* sense and without moral judgement attached (e.g. we can discriminate between blue and green). However, we believe that to discriminate against someone is to wrong them in making this distinction. If so, we use "discrimination" in a *normative* sense. Alt-right and men's activist groups use this ambiguity between descriptive and normative discrimination to their advantage by claiming that any and all distinction making is discrimination and

wrong. Dubbing affirmative action programs, "reverse discrimination" therefore implies that they are wrongful. In response to this, defenders of affirmative action have taken to calling it "positive" or "compensatory discrimination". This points at the strategy to defend affirmative action by claiming that it is instrumentally justified, benign or otherwise allowed discrimination. However, even then, it is still asserted that affirmative action is indeed a case of a *prima facie* morally wrongful act which is made permissible only by circumstance. The negative connotation is not neutralised or removed by this. Rather, it is reaffirmed with every utterance.

In what follows I aim to detangle the descriptive from the normative sense of discrimination and examine the question whether affirmative action is adequately described as discrimination. To this end, I begin by outlining what I understand affirmative action to be. I then introduce two different ways of thinking about the wrong of discrimination, namely harm and social meaning, and apply them to affirmative action.

## 1. Affirmative Action

The first question we need to address is what exactly affirmative action is and how it came to be referred to as "reverse" or "positive" discrimination. "Affirmative action" means positive steps taken to increase the representation of women and minorities in areas of employment, education, and culture from which they have been historically excluded" (Fullinwider 2018, 1). Even more broadly, affirmative action can be implemented for any social group which (previously) faces oppression, discrimination or disadvantage qua social group. Measures that fall under the umbrella of affirmative action include those that ensure equal or proportional representation of disadvantaged groups, e.g. quotas; measures that influence the selection for positions in favour of disadvantaged

<sup>1</sup> By woman<sup>\*</sup> I mean every person that self-identifies as a woman. This obviously includes trans- and cis-women, as well as other gender non-conforming individuals who believe that woman<sup>\*</sup> describes them (c.f. Poertzgen, 2019). Conversely, the same is true for men<sup>\*</sup>. I am aware that in postulating this I am disregarding potential additional discrimination faced by non cis-women and non cis-men.

groups, e.g. point systems; measures to increase applications of members of disadvantaged groups, e.g. targeted advertisement for open positions; and measures to enable persons belonging to disadvantaged groups receive the relevant qualifications, e.g. support networks, mentoring programs, targeted scholarships. All these measures involve distinguishing between persons based on a trait they presumably have or lack, such as belonging to a certain social group (i.e. having a certain gender or race). Thus, on the face of it, affirmative action looks a lot like discrimination. The condition both discrimination and affirmative action share is:

- (1) A is selected in act  $p$  because of some trait they have or lack.

However, as we will see, there is more to discrimination, which makes it a difficult concept to grasp. Calling affirmative action reverse discrimination then, implies that it is wrong. In other words, there seems to be the implication that there is something morally wrong with the distinction making involved in affirmative action measures. For instance, one may argue that it wrongs those individuals belonging to the group that is not singled out for affirmative action by violating their rights (Thomson 2002; Simon 2002). That affirmative action wrongfully discriminates is the charge that arguments for affirmative action usually address. Hence, the question being addressed is whether discrimination for the sake of gender or racial equity<sup>2</sup> is justified or permissible. However, the question of whether affirmative action should indeed be described as an instance of discrimination is rarely, if ever, addressed (Suk 2017).

As I just implied, I will separate the descriptive and normative understandings of discrimination by terminology to avoid confusion. The descriptive version I will call making a *distinction*, *distinguishing* or *differentiating*; only the normative version will be called *discrimination*, *discriminating* or *discriminatory*. This means that there is always and necessarily something morally wrong with the distinction that is made in cases of discrimination—discrimination is always wrongful discrimination. I hence stipulate one necessary condition for discrimination: the act of distinction  $p$  wrongs A. This implies that before we can call some act discrimination, we need to examine whether it does indeed wrong someone.

To return to affirmative action and in order to get a clear-

er understanding of my subject here, I set out a paradigm case for an affirmative action. I take it that quota systems are the most controversial practice out of the affirmative action measures I listed above because under a quota system, the employer is required to hire a person with the selected trait (e.g. women\*) from the pool of applicants. In the following, I will investigate whether such quota systems fulfil the conditions for discrimination, or not. Crucially, I assume that a quota does not require hiring a person that has the desired trait if they do not fulfil the necessary qualifications. Thus, a quota system would not result in hiring unqualified people.

Let us assume the following case: the philosophy institute at our university introduces a gender-quota of 50% among professors, lecturers, post-docs and PhD students. Currently, between 0% and 35% of these positions are occupied by women\*, and the percentages decrease with seniority (Paxton, Figidor, and Tiberius 2012). With the next opening of a position it is clear, that (if possible) a woman\* should be employed. As a position opens, Alice, a woman\*, and Adam, a man\*, are among the applicants. Both fulfil the application criteria. However, Adam has published more frequently (which is in line with findings concerning the representation of women\* in top-tier journals, see Haslanger, 2008). Nevertheless, since Alice has all relevant qualifications and fulfils all other criteria set out in the advertisement, the university decides to hire Alice to achieve the gender quota. I purposefully designed this example in a way that is challenging to other accounts defending affirmative action (Thomson 2002; Dworkin 2002; Cohen 2002) because Adam is more qualified in the example according to traditional understandings of qualification, which seems to further the intuition in people that there is something morally wrong with hiring Alice and, by extension, quota systems.

I turn now to deconstructing normatively salient concepts connected to affirmative action—making a distinction based on traits; making a distinction based on traits related to group membership; preferences and preferential treatment—and to analysing how they are related to discrimination. In the process, I hope to clear up any confused intuitions about affirmative action, as well as distil further necessary conditions for discrimination.

Affirmative action involves distinguishing between persons on the basis of some characteristic they presumably have or lack, which marks them as belonging to a certain social group. One might argue that this is in and of itself morally problematic (c.f. Hellman, 2017, 98). One could argue for instance that selecting Alice based on her trait, i.e. being a woman\*, is doing her an injustice by not treating her as an individual. In this view, there is something intrin-

---

<sup>2</sup> I use “equity” rather than “equality” because it implies an understanding of substantive equality. In order to achieve equality, we need to treat people according to their needs and not just strictly the same.

sically wrong with condition (1). While it may be the case that sometimes selecting someone based on a trait they have or lack is wrong, there are plenty instances in which this is not the case. It might be morally wrong at times not to distinguish based on traits signifying group membership. There are, arguably, good reasons for specialized medical professionals such as gynecologists. They have good reason to only treat women\* and trans-men. Thus, it cannot be the case that making distinctions between persons based on some trait they have or lack is in itself morally wrong.

Alternatively, one could make the point that what makes making a distinction based on traits a person has or lacks morally wrong is that these traits are morally arbitrary and should therefore be irrelevant for the decision. These morally arbitrary traits include, but are not limited to, gender and race<sup>3</sup>. However, if this were the case, then it would follow that providing particular drugs for African-Americans or women\* would be wrong (example in Hellman, 2008, 5). If we suppose that what makes a distinction involving race or gender wrong is that these are irrelevant traits, then we have to face the problem that a trait's relevance depends very much on the context of the decision. Hence, it might be that an employer argues that hiring women\*, who might become pregnant or have children and are thus presumably available only for certain schedules, decrease efficiency and productivity of the company. The "relevance" of the trait (i.e. gender) in this case results from its correlation with the criterion of efficiency (also in Hellman, 2008, 6). Thus, a distinction is not *pro tanto* wrong because it is based on the traits of a person that link her to a social group, even when these are morally arbitrary or irrelevant.

I have just argued that (1) does not necessitate the conclusion that A is wronged by *p*. Therefore, we need at least one additional condition. Let me summarize the necessary conditions for discrimination we have thus far identified:

- (1) A is selected in act *p* because of some trait they have or lack.
- (2) Thereby, A is wronged by *p*.

Of course, act *p* may be justified or excused by an overriding consideration. These might be other obligations, rights or considerations of the overall good. Hence, an act of discrimination *p* may be, all things considered, right if other reasons for the action are considered (Arneson 2017, 152). This is the thrust most defences of affirmative action take.

I mention it here to avoid confusion, but I will not further elaborate on this possibility beyond assuming it exists. The conclusion that there is more to discrimination than the fact that certain distinction making is always wrong, is one most scholars can agree upon (Hellman 2017; Arneson 2017; Lippert-Rasmussen 2014). However, there are many different accounts of what it is exactly that makes some distinctions wrongful and discriminatory. In the following, I discuss some very common ideas as to why certain distinctions are wrong that are linked to affirmative action. The first intuition many share regarding affirmative action is that it is understood to involve differential treatment, such as preferences, and that this is problematic. Genderquotas for instance give preference to women\* over men\*. In everyday language use, giving preference refers to choice rankings, prioritizations and comparative evaluations. In philosophy, we can conceptualize preferences "as subjective comparative evaluations" (Hansson and Grüne-Yanoff 2017). Preferences *per se* do not seem to be morally problematic to us. My preference for apples over pears is no issue for normative argument. My preference for spending time with my family rather than strangers is similarly unproblematic. What seems to make preferences morally problematic is the different treatment it may result in. Accordingly, preferential treatment implies that someone is treated positively different than others. What is problematic about the distinction making in question is therefore its consequences.

Is this what causes the intuition that introducing a gender-quota of 50% in the philosophy institute of your university is morally problematic? With every position that opens after the quota was established, we understand that the university has made the evaluation that it would be better to hire women\* as opposed to men\*. Hence, women\* would receive favourable treatment with a quota in place. They have an advantage; therefore, preferential treatment is in violation of the principle of equality—or so the argument goes. However, preferential treatment is also widely accepted as morally unproblematic. Awarding positions based on qualification or ability to pass relevant tests is common practice and seems unproblematic to most of us. To practice medicine, you need the relevant qualifications; to drive a car, you need to pass a test and be of a certain age. Similarly, my preference for helping my close friend move house rather than help an acquaintance seems untroubling to most of us (all things being equal that is). Hence, that an act involves preferential treatment is not a necessary nor sufficient condition for discrimination.

Rather than looking at the preferential treatment women\* receive through a quota, one may focus on the comparative disadvantage this places men\* in. Accordingly, affirmative

---

<sup>3</sup> When I use the term race, I refer to the social construct that racializes certain people. I take as given that biologically, race does not exist.

action is problematic and discriminatory because it puts people at a comparative disadvantage. However, being put at a competitive advantage is not yet sufficient to make a distinction wrong. In this I take this case to be analogous to being preferentially treated. One could imagine a sports coach who puts their best player at a disadvantage in order to benefit her overall by making her train harder.<sup>4</sup>

Consequently, distinguishing between people based on traits they have or lack is not a sufficient condition for wrongdoing. This also means that affirmative action is not discrimination only because it involves an act of distinguishing between persons based on traits. Additionally, that the choice has positive consequences for, in our case, women\* is not sufficient to establish that it is wrongful. Similarly, that it puts someone at competitive disadvantage is not sufficient to make a distinction wrong. This, however, leaves us with a second condition for discrimination and affirmative action. So far what is clear is that an act  $p$  is discriminatory only if:

- (1) A is selected in act  $p$  because of some trait they have or lack.
- (2) Because of this, A is treated differently.

That discrimination and affirmative action share these conditions does not yet allow the conclusion that affirmative action is morally wrong. Moreover, something additional is needed to explain what makes some cases of drawing distinctions wrongful and, by stipulation, discrimination. In order for affirmative action to be discrimination, this element would also have to be present in the quota case I described above. Wrongful discrimination needs to fulfil these additional conditions:

- (3) ?
- (4) Thereby, A is wronged by  $p$ .

Whether or not affirmative action also fulfils them, is yet to be determined.

In the following, I examine two accounts of the wrong of discrimination, which offer ways to fill in condition (3). The first, I think, is a very common assumption about discrimination, i.e. discrimination harms. The second account of discrimination is one that appears more plausible to me, i.e. discrimination demeans. I will begin with the former, which I call the harm-view of discrimination. Under this view we ask whether the person not getting the job due to an affirmative action program or, in our case, Adam, has been harmed and thus wronged because of the quota.

---

<sup>4</sup> Thanks to Kelly Tuke for this example.

## 2. The Harm-View of Discrimination

The harm view of discrimination holds that discrimination is wrong because it harms someone. Thus, an act of distinction-making  $p$  is discrimination and *pro tanto* wrong iff

- (1) A is selected in act  $p$  because of some trait they have or lack.
- (2) Because of this, A is treated differently; and
- (3) A is harmed by  $p$ .
- (4) Therefore, A is wronged by  $p$ .

In our example Adam is selected in the act  $p$  because he lacks the trait “woman”. As a consequence, he is not hired for the open position. The question is: has Adam been harmed and therefore been wronged by act  $p$ ? In other words: has Adam been discriminated against? For the sake of clarity, I will also turn this case on its head and consider whether Alice has been wronged by being chosen because she is a woman\*.

In order to answer this question, we need to clarify what we mean by harm. Several different views on this exist in the literature. First, there are comparative views (Arneson 2017). One version of this view holds that to harm somebody is to make them worse-off. This, however, may again be interpreted in different ways: firstly, one may be made worse-off in a counterfactual sense. Thus, A is harmed iff A is made worse-off by  $p$  than A would have been had  $p$  not been performed (see Arneson, 2017). This view covers the idea that Adam has been put at a comparative disadvantage because of the gender quota. Secondly, one may also be made worse-off in a temporal sense: A is harmed iff A is made worse-off by  $p$  than A was before  $p$  was performed (Harman 2004; 2009). Let us consider them individually.

### 2.1 The Counterfactual View

First, let us look at the counterfactual view. Under this view an act counts as discrimination iff

- (1) A is selected in act  $p$  because of some trait they have or lack.
- (2) Because of this, A is treated differently; and
- (3) A is made worse-off by  $p$  than A would have been had  $p$  not been performed.
- (4) Therefore, A is wronged by  $p$ .

Consider our case in which Adam and Alice both apply for the same position. Adam is slightly more qualified on one metric. Nevertheless, because Alice is a woman\* she is hired under the quota. Alice is selected because of a trait she has—being a woman\*. The question would now be whether she has been harmed by this act. Alice

is comparatively better-off after being hired in act  $p$  than she would have been had act  $p$  not been performed.<sup>5</sup> One could of course argue at this point that Alice has been made worse-off because this job is one that is comparatively more stressful than whatever she would have been doing had she not been hired, and may lead to burn-out for instance. This, however, seems to be a paternalistic argument to make because it assumes we know better than Alice what is good for her. Let us assume, thus, that Alice has considered her options well before applying and that getting the job makes her better-off. Therefore, she has not been discriminated against under my understanding of the term.

Critics may now point out that there is another person affected by the distinction that is being made: Adam. This would point to understanding quotas as reverse discrimination. Whereas previously (in this understanding) women\* have been discriminated against in hiring decisions, it is now men\* who are discriminated against because of gender quotas. The question is thus: has Adam been discriminated by the gender-quota? Or in other words, has Adam been made worse-off than he would have been had the action  $p$  not been performed? If we want to argue that he has, we would commit ourselves to assuming that not performing the action (hiring Alice) would result in Adam being hired. However, not hiring Alice does not yet mean hiring Adam, as there is always the option of not hiring anyone. Thus, he has not been made worse-off than he would have been because had the action not been performed, he would still be in much the same position he is in now—someone looking for employment.

One might object that this is misunderstanding the counterfactual view, that Adam has been made worse-off because the alternative action to hiring Alice was to hire him. However, this view would result in harm being unavoidably linked to preferences. We would not be able to maintain then that having preferences of any kind regarding persons and acting on them would be permissible. In deciding, for instance, whether to have dinner with one's family or with a friend, we would end up inadvertently harming one of the parties. Similarly, a father would be harming all the children on the playground that are not his to take care of, by not giving them the same attention he gives his own child. This judgement however, appears rather counterintuitive to us. Of course, one can have the view that there can be situations in which either decision ends up being wrong. However, I would not want to argue that making a decision between persons based on some preference is always a moral dilemma. This is because I do

not view these types of choices as harming, but of benefitting someone.

Under this understanding, Adam is not being benefitted by not being hired. Arguably, there is no moral obligation to benefit anyone (Thomson, 2002). The situation presents itself very differently this way, than if we assume that not hiring someone is harming them. Similarly, withholding a benefit is not *per se* harmful. If it were, we would be morally obligated to benefit all kinds of persons all the time. There may be situations in which we ought to benefit a person, I claim here though that awarding a person a job is not one of these situations.

## 2.2 Temporal view of harm

There is another way of understanding what it means to harm a person. This I call the temporal view. It holds that we harm a person if they are made worse-off by the act than they were before it was performed. Accordingly, act  $p$  is discrimination iff:

- (1) A is selected in act  $p$  because of some trait they have or lack.
- (2) Because of this, A is treated differently; and
- (3) A is made worse-off by  $p$  than A was before  $p$  was performed.
- (4) Therefore, A is wronged by  $p$ .

Has Alice been discriminated against based on the temporal view? Arguing that Alice is worse-off after being hired would involve a pretty convoluted argument, but it is not impossible. However, let us assume that she knows what she wants and what she is getting by applying for the job. Under normal circumstances it is quite obviously not the case that she is worse-off now than she was before.

But what about Adam? Is he made worse-off than he was before the action? One could argue that he might be worse-off psychologically. Maybe this was the tenth time he was not considered for a job. However, this would make harm a subjective notion: A person would be harmed if she feels that she is being harmed. I do not mean to argue that these feelings are irrelevant, nor that this is impossible. However, I believe that it would lead to many potentially counter-intuitive moral judgements. Think again of my friend who I did not have dinner with because I had dinner with my family instead. She could now claim to have been harmed by me. By not having dinner with her, I have wronged her. The moral permissibility of making the choice to eat with my family now depends on my friend's perception of being harmed. I would think that this appears counter-intuitive to most people. Moreover, arguing that not hiring a person is in and of itself sufficient for

---

<sup>5</sup> This is assuming that she wanted and maybe needed the job.

harming them would normatively undermine our hiring processes and I am quite sure that most people would agree that it makes little sense. Thus, arguing that Adam has been made worse-off by the decision to hire Alice is not convincing.

However, what I believe becomes even clearer now, is that the question of getting a job is in fact a question of being benefitted or not being benefitted. For Adam, not getting the job means that he does not receive the benefits which come with it, however, as I have argued above, equating not being benefitted with being harmed would stretch the boundaries of our moral theories.

### 2.3 Non-comparative harm

All I have said so far might be a reason to change our view of harm to a non-comparative account. Elizabeth Harman proposes that one can be harmed even though one is not over-all made worse-off. She argues that one has been harmed if one is in a condition that “causes pain, early death, bodily damage, or deformity” (Harman 2004, 93). On this view however, discrimination would be limited to cases that negatively impact a person’s physical health. Neither Adam nor Alice can plausibly claim to have been harmed in this manner by the decision to give the job to Alice.

Considering whether Adam or Alice have been harmed by the decision to hire Alice also indicates that the harm account of discrimination is lacking plausibility. To illustrate the problems, imagine another case in which a Black woman\* is not hired because of her race and gender. Has she been made worse-off? We would have to say that she has not. We could even imagine a scenario in which she is benefitted by this decision as she got a much better job offer afterwards (see Arneson 2017, 159–160). Regardless of this, I believe that most of us still think that she has been treated wrongfully. There is something wrong with denying her because of her race and gender. We could now argue that not receiving the job due to racial and sex-based prejudice can make her worse-off in a psychological sense, but that would mean that our account of harm is subjective yet again. However, we could also imagine a scenario in which she never learns about the reasons why she did not get the job. I think most of us still have the intuition that she has been discriminated against. That the harm account of discrimination, in the interpretations that I have discussed so far, cannot account for these cases should “cast further doubt on the claim that . . . discrimination must be harmful to qualify as morally wrong” (Arneson 2017, 159).

Hence, I conclude that firstly, under a harm-view of discrimination, the quota system, as the most controversial case of affirmative action, is not discrimination. The decision to hire Alice over Adam does not make him worse-off

in any sense of the word, nor does it put him in a harmful state. The same is true for Alice. Secondly, the discussion has shown that a harm-view of discrimination is not very plausible.<sup>6</sup> Let us hence turn to another conception of discrimination—the social meaning-view.

### 3. The Social Meaning-View of Discrimination

Deborah Hellman (2008; 2017) argues that an act counts as discrimination (and therefore as wrongful) if the distinction involved demeans, that is, treats the person as less than of equal human worth. It is very important to note here that Hellman allows for distinguishing acts to be wrong independently of the wrong that makes them discrimination.<sup>7</sup> Thus, an act *p* is discrimination iff:

- (1) A is selected in act *p* because of some trait they have or lack.
- (2) Because of this, A is treated differently; and
- (3) A is demeaned by *p*.
- (4) Therefore, A is wronged by *p*.

What does it mean for an act to be demeaning? Hellman states that she understands an act to be demeaning if it debases or degrades or in another way betrays “disrespect for the equal humanity of the other” (2008, 35). A demeaning act is one that attempts to lower the social status of the person(s) affected.

That an act demeans is therefore a claim about the objective *meaning* of the act (Hellman 2017). This implies that neither intention or motive, nor consequences or results of the act are important to its normative status. Additionally, neither the meaning derived from the intention of the agent, nor the interpretation of the person affected is what determines if an act demeans. Rather, that an act has this meaning depends on the context of the act and on what it therefore expresses (25–27). Hellman includes speech acts in this, with the important consequence that statements may discriminate (38). Hellman’s account of discrimination is a particularistic one: for each action that distinguishes among people we need to investigate whether the context is such that the act is demeaning. This also means that the objective meaning of an act can be a matter of debate. However, Hellman states that for some actions it is very clearly the case that they demean. As an example

<sup>6</sup> For a more thorough discussion of the harm view of discrimination and other arguments against its plausibility see Arneson (2017).

<sup>7</sup> Hellman does not distinguish between descriptive and normative discrimination in the same way I do. Thus, for her it is possible that an act is discrimination without it being the case that it is wrong by being demeaning, or an act may be discrimination and be wrong, but not for the reason that would make it wrongful discrimination in her account (see pp. 13–21).

she proposes the Jim Crow laws of the US-American south (Hellman 2017). Whether an act is demeaning hence depends on the social context of the action in question. The relevant social context of an action can be both historical and current (2008, 21). The questions we need to ask ourselves are: Has the trait in question been used historically to demean or oppress a group of people? Are people who share this trait socially situated in a way that is indicative of lower social status?

Selecting people for differential treatment based upon their skin colour, for instance, may be demeaning in a society with racist history and structures. A teacher asking students to be seated according to their skin colour would thus wrongfully discriminate, but we would not have the same intuition if she asked her students to sit according to the colour of their clothes. This is because skin colour did and does make a difference for people's lives and implies racialized oppression based on an ideology of white supremacy, while the colour of clothes does not have the same implication, as far as I am aware. That the meaning of the act is context dependent, however, allows for this situation to be interpreted differently as well: imagine the teacher asked her students to sit according to their skin colour in order to demonstrate apartheid or Jim Crow. This may be sufficient to neutralize the situation and make it so that it is not demeaning (28). What matters is thus firstly, "how that characteristic has been used to separate people in the past and the relative social status of the group defined by the characteristic today" (28), and secondly, whether the resulting treatment "is better or worse in a way that would be true across cultures... Other times, the difference in treatment resides in the cultural significance of that treatment" (28). Hellman's account thus is one in which the act of distinguishing is made wrong by the meaning of the treatment. Consequently, not every act that is disrespectful is also demeaning. Traits that have historical or current social significance have more potential to be the basis for a demeaning distinction, or in other words, to be discriminatory (41).

There is another dimension that plays into whether an act can be considered demeaning, which is whether or not the agent has *power* relative to the person they affect (35). Without this relative power an act may insult but not demean (36). A little boy ordering a woman does not have the relevant power to demean. This is because he lacks the authority necessary to put the other down or to (threaten to) lower their relative status (38; 2017, 103); Hellman argues that to demean is like ordering in this sense. However, it need not be the case that the demeaning act in fact lowers the status of the affected person—another similarity to ordering. What is relevant is not the intention or

the consequences but the meaning of the act itself. As a consequence, acts can be demeaning even if no one knows about them (2017, 105). What matters to whether an act is demeaning or not is whether or not the actor and the act have the *capacity* to demean.

Hence, whether a distinction is demeaning and thereby discriminatory is a question of both an expressive and a power dimension. Consequently, private actions are less likely to be demeaning than governmental actions because the agents in question lack power and the same is true for individual as opposed to institutional acts (2008, 57).

Let us return now to the question of whether Adam or Alice are being demeaned by the gender quota. Hellman herself discusses affirmative action in connection with discrimination. She argues that whether affirmative action should be considered discriminatory depends on whether or not the specific act in question expresses that the group with that trait, or persons in that group, have lower status. Her case is one of a quota aimed at increasing diversity, in which a white person does not get hired. Hellman argues that while the act may have other negative consequences for the person who is not being hired, there is nothing in the act that expresses that whites are of lower status. The quota is therefore not demeaning to white people (2017, 101). If we assume that the case is parallel to that of gender equity, as per our example, Adam is not being demeaned, because the act does in no way express that he is of less than equal moral worth.

Before drawing this conclusion, I think it is necessary to investigate in which senses a gender quota could be considered demeaning for Adam. One could imagine that Adam argues that he is not treated as an individual but rather as a kind, namely man\*, and that this expresses disrespect for his equal moral worth. The first step is to argue that in selecting Alice, Adam has not been treated as an individual. This is somewhat empirical and in and of itself not sufficient to establish that he has been wronged (as I discussed in section 2). The second step is to argue that not being treated as an individual is demeaning, which would establish that he has been wronged. Why would this be the case? One could argue that focusing only on one trait fails to afford Adam the consideration he deserves as a human being. By reducing him to the trait "man" one disregards his multifaceted human nature. Arguably though, the policy does not deny that there is more to Adam than that he identifies as a man\*. Its focus on the trait of gender is a question of efficiency of administering support to a previously disadvantaged group and not one of reducing persons to one of their many traits. Adam may therefore well be insulted that the policy happens to focus only on one trait, but he cannot convincingly argue that this fact is

threatening or lowering his status as a human being. If policies that focus on one trait only were thereby demeaning, we would have a problem regarding all policies that focus on groups via a trait they have or lack. For instance, a child could complain that not allowing her to drive a car fails to treat her as an individual with individual capacities and abilities and therefore demeans her. Importantly, Alice would also have possible grounds for complaint against the gender quota. After all she has been selected because of a trait of hers.

Another possible response would be that a gender quota is not demeaning for men\*, as it is not being addressed at them, but at women\*. As such, one could argue, the policy has nothing to say on the relative moral worth of men\*. However, as Hellman states, “[t]he claim that blacks, rather than whites, are insulted by affirmative action policies is more plausible.” (101) The idea is that affirmative action programs suggest inferiority in terms of ability and intelligence for those who benefit from the quota being in place. If this were the case, quotas aimed at gender or racial equity would be demeaning to those who are addressed by it. Is this a plausible interpretation? I would argue that it is not, because it presupposes that those hired due to a quota are unqualified, which makes no sense as a general statement and is not applicable in the scenario under consideration. This reading is plausible only if the policy was to hire women\* regardless of their qualification. However, as I have established, both Adam and Alice are qualified. The policy is to hire qualified women\*, if possible. Alice received the job both because of her qualifications and the fact that she has the trait woman\*. Thus, the policy does not express inferiority in terms of capability. As Justice Stevens has argued, preferences for minorities should not be misunderstood in this way and that to do so is to confuse “a ‘No trespassing’ sign and a welcome mat” (*Adarand v. Pena* 1995, 245 in Hellman, 2017, 102). One could alternatively focus on the aspect of help rather than ability and intelligence. Could the quota be demeaning to Alice because it implies that she needs help to get to the position? Is needing help an expression of one’s lesser moral worth? I would hope not. If it were, then our moral worth would depend upon our capabilities. I believe we can agree that the moral worth of a person does not depend upon what she can or cannot do. Thus, our gender quota does not demean Alice.

In fact, one could argue the opposite: Not establishing a gender quota would very likely be demeaning to Alice. Let me explain this. If the university does not hire Alice because she is a woman\* this would clearly be discriminatory. Without the quota in place, Alice would most likely not be hired due to being less qualified. This is not

as obviously demeaning and therefore discriminatory as not hiring her because she is a woman\*. As I have established, the only difference between Alice and Adam can be construed in terms of merit or qualification: Adam has published more frequently than Alice. However, there is evidence to suggest that the difference in publication records may very well be due to bias in the publishing houses (Haslanger 2008). Concluding from the fact that Alice has published less often (or in lesser-indexed journals) that she is less qualified is therefore not straightforwardly possible. If Alice is less qualified because she has published less, and she has published less because she is a woman\*, this turns a decision to not hire her because of her publication record into a proxy for not hiring her because of her gender.<sup>8</sup>

What conclusion should we draw from this problem? I would argue that it is very likely that a decision to not hire Alice in the situation I have described is discriminatory. This is because not hiring her in the face of the historical and current oppression and discrimination of women\* expresses a lack of consideration for her situation that could be considered demeaning. Not paying these circumstances the attention they command by virtue of a concern for equality could thus well be interpreted as demeaning to Alice because it implies that the plight of women\* (and thus Alice) is of little relevance, which betrays a lack of concern for equality and thus for women\*s moral worth. The university does not care or acknowledge that she has (most likely) faced challenges that Adam has (most likely) not faced.<sup>9</sup> In fact, by not hiring Alice because of her publication record, the university builds on and contributes to the structural disadvantages Alice has faced. Not establishing the gender quota is equivalent to the university saying: we do not care about the oppression women\* face in academia. One could argue therefore that the gender quota prevents possible further discrimination as would have likely happened to Alice had the quota not been in place.

I need to address the obvious objection that in Alice’s case none of this is true. We cannot be sure that it was the case for Alice that she has not been published because of her assumed gender. It may well be the case that her writing is not as good as Adams. In this case a decision

<sup>8</sup> This points to a general problem with our conception of qualification and merit. How do we judge who is (more) qualified? Which standards are reliable? How do we make sure that they do not obscure underlying structural disadvantages faced by a group? Addressing these questions would require more room than I have in this essay.

<sup>9</sup> I say “most likely” because there is still the possibility that Adam might be disadvantaged due to other factors. I by no means want to argue that women\* always are in an inferior social position than men\*.

not to hire her would not be demeaning to her. I would still maintain though that a gender quota makes sense if only to prevent the possibility of discrimination. As I have shown, Adam has little to no ground for complaint with a gender quota in place. Therefore, having one may be the least bad option available to us.

## Conclusion

In sum, affirmative action is not discrimination according to the harm-view of discrimination nor according to the social meaning view of discrimination. Rather, as I have argued, affirmative action measures, such as a gender quota, may be useful to prevent further structural discrimination. This alone, however, is only one way of arguing in favour of establishing quotas and other measures.

Opponents of affirmative action could point to potential negative consequences of establishing such measures. For instance, the assumption that people hired due to a quota lack qualification might lead to epistemic injustices, which would need addressing if this were the case. The question is whether the possibility of affirmative action creating epistemic injustices is a sufficient reason not to engage in it. Also, it is questionable whether this is not part of the overarching credibility deficit women\* and other minorities already have (c.f. Fricker, 2007). The assumption clearly seems to be that they (women\* and other minorities) are less capable. Further research would be needed to engage in some of the problematic implications of a quota and into whether this out-weighs the potential benefits of one.

Thus, I do not mean to claim that affirmative action is the perfect or even a good solution for achieving gender or racial equity or even for challenging the current social order of white supremacist patriarchy. It is, however, a solution available to us, one that has the potential of challenging gendered and racialized norms in the job-market and therefore has to be considered. What it cannot achieve is the erosion of socio-economic privilege. Affirmative action alone will also not bring about substantive equity across dimensions of gender, race, age, able-bodiedness, sexuality, socio-economic standing and other intersections –this would require more substantive changes.

## References

### Journal Articles and Books

- Arneson, Richard. 2017. "Discrimination and Harm." In *Routledge Handbook of the Ethics of Discrimination*, edited by Kasper Lippert-Rasmussen, 151–64. Routledge.
- Cohen, Carl. 2002. "Who Are Equals?" In *The Affirmative Action Debate*, edited by Steven M. Cahn, 95–102. New York: Routledge. <http://www.debate.org/affirmative-action-debate/>.
- Dworkin, Ronald. 2002. "Bakke's Case: Are Quotas Unfair?" In *The Affirmative Action Debate*, edited by Steven M. Cahn, 2nd ed., 103–12. New York & London: Routledge. <http://www.debate.org/affirmative-action-debate/>.

- Fricker, Miranda. 2007. *Epistemic Injustice - Power & the Ethics of Knowing*. Oxford: Oxford University Press.
- Fullinwider, Robert. 2018. "Affirmative Action." In *Stanford Encyclopedia of Philosophy*. <https://doi.org/10.5860/choice.41sup-0181>.
- Hansson, Sven Ove, and Till Grüne-Yanoff. 2017. "Preferences." In *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/preferences/>.
- Harman, Elizabeth. 2004. "Can We Harm and Benefit in Creating?" *Nous-Supplement: Philosophical Perspectives* 18 (1): 89–113. <https://doi.org/10.1111/j.1520-8583.2004.00022.x>.
- . 2009. "Harming Future Persons." *Harming Future Persons: Ethics, Genetics, and the Nonidentity Problem*, 137–54. <https://doi.org/10.1007/978-1-4020-5697-0>.
- Haslanger, Sally. 2008. "Changing the Ideology and Culture of Philosophy: Not by Reason (Alone)." *Hypatia: A Journal of Feminist Philosophy* 23 (2): 210–23. <https://doi.org/10.2979/hyp.2008.23.2.210>.
- Hellman, Deborah. 2008. *When Is Discrimination Wrong?* Harvard University Press.
- . 2017. "Discrimination and Social Meaning." In *Routledge Handbook of the Ethics of Discrimination*, edited by Kasper Lippert-Rasmussen, 97–107. Routledge.
- Lippert-Rasmussen, Kasper. 2014. "Racial Profiling." In *Born Free and Equal - A Philosophical Inquiry into the Nature of Discrimination*. New York: Oxford University Press. <https://doi.org/10.1017/CBO9781107415324.004>.
- Paxton, Molly, Carrie Figdor, and Valerie Tiberius. 2012. "Quantifying the Gender Gap: An Empirical Study of the Underrepresentation of Women in Philosophy." *Hypatia* 27 (4): 949–57. <https://doi.org/10.1111/j.1527-2001.2012.01306.x>.
- Poertzgen, Karen. 2019. "Giving up the Gender Debate." *Meta(φ)* 5. [https://www.fsphilosophie.unibe.ch/unibe/portal/fak\\_histisch/dkk/philosophie/fachschaft\\_philosophie/content/e557932/e778728/e778737/e885354/e885356/Metaphi\\_N5\\_ger.pdf](https://www.fsphilosophie.unibe.ch/unibe/portal/fak_histisch/dkk/philosophie/fachschaft_philosophie/content/e557932/e778728/e778737/e885354/e885356/Metaphi_N5_ger.pdf).
- Simon, Robert. 2002. "Preferential Hiring: A Reply to Judith Jarvis Thomson." In *The Affirmative Action Debate*, edited by Steven M. Cahn. New York & London: Routledge. <http://www.debate.org/affirmative-action-debate/>.
- Suk, Julie. 2017. "Discrimination and Affirmative Action." In *Routledge Handbook of the Ethics of Discrimination*.
- Thomson, Judith Jarvis. 2002. "Preferential Hiring." In *The Affirmative Action Debate*, edited by Steven M. Cahn. New York & London: Routledge. <http://www.debate.org/affirmative-action-debate/>.

### Newspaper Articles

- Binkley, Collin. 2018. "Harvard, NYU law reviews accused of bias against whites, men." *AP, October 9, 2018*. <https://apnews.com/c41c7ec12fa44d3bac4106605a47f3e5>.
- Lefkovitz, Alison. 2018. "Jordan Peterson and the return of the men's rights movement." *The Washington Post, July 24, 2018*. <https://www.washingtonpost.com/news/made-by-history/wp/2018/07/24/before-jordan-peterson-there-were-mens-rights-activists/>.

---

**Karen Poertzgen** interessiert sich besonders für intersektionale feministische und anti-rassistische Theorie und angewandte Ethik. Karen hat kürzlich ein Doktorat mit dem Thema "Individuelle moralische Verantwortung für strukturelle Ungerechtigkeiten" hier in Bern begonnen.

# Wittgensteins Parallelen zu der buddhistischen Realitätsvorstellung

## Einleitung

„Erwach, erwach doch  
Und werde mein Gefährte  
Verträumter Falter!“  
Bashō<sup>1</sup>

Wittgenstein erwähnt die Begriffe *Welt* und die *logische Form der Wirklichkeit* mehrmals im *Tractatus (TLP)*,<sup>2</sup> ohne das Verhältnis der beiden Begriffe deutlich zu erklären. Ein möglicher Ansatz zum Verständnis dieser beiden Begriffe besteht darin, dass die Welt beschreibbar ist. Sie ist die Summe aller Tatsachen (TLP, 1.1). Diese Tatsachen lassen sich anhand von Aussagesätzen erfassen und dadurch beschreiben. Die Aussagesätze reichen so zur Wirklichkeit hin (TLP, 2.1511). Die logische Form der Wirklichkeit lässt sich hingegen nur zeigen (TLP, 4.12 und Vossenkuhl 2001, 56). Diese Ausführungen erzeugen die Vermutung, dass sich Welt und die logische Form der Wirklichkeit voneinander unterscheiden. Welche begriffliche und metaphysische Implikationen entstehen aber aus dieser Unterscheidung? Diese zwei Begriffe zu klären und in ein Verhältnis zu bringen, wird die Arbeit strukturieren.

Die vorliegende Untersuchung geht dabei von der folgenden Arbeitshypothese aus:

Man kann deutliche Parallelen zwischen Wittgensteins Begriffen *Welt* und *logische Form der Wirklichkeit* und der buddhistischen Realitätsvorstellung finden. Diese Parallelen zeigen sich auf der epistemologischen, der ontologischen und der normativen Ebene.

Viele buddhistische Strömungen führen eine Unterscheidung ein zwischen der konventionalen Realität

(*samvrti satya*) und der ultimativen Realität (*paramartha satya*) (Mortensen in D'Amato et al. 2009, 3). Diese Realitätsvorstellung ist nach vielen buddhistischen Strömungen auf der epistemologischen, der ontologischen und der normativen Ebene je als unterschiedlich zu behandeln und zu verstehen. Alle drei Ebenen hängen eng miteinander zusammen, was sich noch deutlich zeigen wird. Die vorliegende Arbeit versucht anhand dieser drei Ebenen, die Begriffe *Welt* und *logische Form der Wirklichkeit* von Wittgenstein besser verstehen zu können. Es werden dabei alle drei Ebenen betrachtet, um zeigen zu können, dass die Ähnlichkeiten nicht nur willkürlich aufgegriffen werden, sondern eine deutliche, strukturelle Parallele besteht. Hierzu wird zunächst im Kapitel (2) die exegetische Methode vorgestellt, welche in dieser Untersuchung zur Anwendung kommt. Sie wird hier *metaphorischer Inferentialismus* genannt. Die folgenden Kapitel (3, 4 und 5) werden jene drei Ebenen der buddhistischen Realitätsvorstellung erläutern und bringen sie mit Wittgensteins Begriffen in Verbindung. Es wird sich zeigen, dass die oben genannte Arbeitshypothese bestätigt werden kann und sich Parallelen auf allen drei Ebenen finden lassen. Im Kapitel (6) wird mit Hilfe des Zen Buddhismus gezeigt, dass die hier erstellte Dualität jedoch wieder überwunden werden muss. Das Kapitel (7) schliesst die Arbeit mit einigen Bemerkungen zu Wittgenstein und dem Buddhismus als Praxis ab. Bevor die Untersuchung nun beginnt, bedarf es im Voraus noch einiger Erläuterungen. Erstens kann die vorliegende Arbeit weder als Einführung in Wittgensteins Tractatus noch in den Buddhismus allgemein verstanden werden.<sup>3</sup> Hierbei werden also einige Punkte vorausgesetzt während andere ausgelassen werden. Dies führt auch gleich zum zweiten Punkt: Buddhistische Strömungen unterscheiden sich deutlich voneinander.

1 In Ulenbrook 2010, 79.

2 Vgl. jedoch TLP, 2.063, 2.18, 4.05f.

3 Vgl. jedoch hierzu Han 2017, Kenny 1974, Priest 2018, Von Brück 2016, sowie Vossenkuhl 2001.

Die hier vorliegende Arbeit bezieht sich daher vorrangig auf den Zen (*Chan*) Buddhismus und es wird deutlich hervorgehoben, wenn von einer anderen Strömung ausgegangen wird. Drittens stellt sich die Frage, welchen Wert eine solche Untersuchung haben kann. Die hier vorliegende Arbeit schliesst sich an einem Projekt an, welches von Philosophen wie beispielsweise Gudmundsen (1986), Read (bspw. in D'Amato et al. 2009) und Priest (2018) durchgeführt wird. In diesen Projekten wird versucht, das Verständnis von Wittgensteins Werk zu verbessern, indem man sich nicht auf Wittgensteins Tradition (Schopenhauer, Kierkegaard, Frege, Russell etc.) beruft, sondern auf Philosophien, zu denen Wittgenstein wahrscheinlich kaum oder gar keinen Bezug hatte. Aus solchen Vergleichen können interessante Einsichten gewonnen werden. Hierbei spielt der Buddhismus eine zentrale Rolle. Dieser Interpretationsansatz ist dementsprechend kritisierbar, was hier jedoch nicht weiter besprochen wird. Die Arbeitshypothese leistet aber einen Beitrag zu diesen Projekten und soll im besten Fall dazu führen, Wittgensteins undurchsichtiges Werk besser verstehen zu können.

## 1. Exegetische Vorannahmen

„Dieses Buch wird vielleicht nur der verstehen, der die Gedanken, die darin ausgedrückt sind - oder doch ähnliche Gedanken - schon selbst einmal gedacht hat.“<sup>4</sup>

Vor der inhaltlichen Auseinandersetzung mit der vorgestellten Arbeitshypothese, müssen hier zuerst die methodischen Grundlagen dieser Arbeit genauer erläutert werden. Die hier vorgestellte Methode wird für die gesamte Arbeit tragend sein. Wenn man sich mit einem klassischen Werk befasst, sei dies nun Wittgensteins Tractatus oder ein buddhistischer Text, so stellt sich die Frage, wie man diesen verstehen kann und auch verstehen will. Es gibt drei mögliche Perspektiven, wie man sich Primärtexten und Kommentaren annähern kann. Diese drei Perspektiven sollen hier *de dicto*, *de re* und *de traditione* Analysen genannt werden. Alle diese drei Perspektiven können als gleichwertig verstanden werden, was bedeutet, dass keine Perspektive der anderen notwendigerweise überlegen ist (Brandom 2002, 104). Die Art wie man sich auf die verschiedenen Perspektiven fokussiert, bestimmt die hermeneutische Methode für die jeweilige Auslegung. Daher sollen die drei Perspektiven hier kurz erläutert werden. Anschliessend wird der *metaphorische Inferentialismus* als die für die vorliegende

Untersuchung gewählte Methode vorgestellt. Einen Text *de dicto* auszulegen, bedeutet, dass man versucht, den konzeptuellen Gehalt der verwendeten Begriffe innerhalb des Textes zu erschliessen (Brandom 2002, 94). Dies ist eine Möglichkeit, die Autor\*innen zu verstehen. Indem man Argumente nachzeichnet, Prämissen hervorhebt und Zusammenhänge erschliesst, wird es möglich, die Bedeutung und den Gebrauch der einzelnen, verwendeten Begriffe freizulegen (Rohbeck in Meyer 2016, 243). So kann man beispielsweise Wittgensteins Auffassung des *logischen Raums* besprechen, oder die Zusammenhänge von Leerheit (*ku*) und sichtbarer Form (*shiki*) im Herz Sutra (*Hannya Shingyo*) untersuchen, indem man sich direkt an die einzelnen Textstellen wendet, wo diese Begriffe auftauchen.

Eine *de re* Analyse stellt hingegen einen anderen Ansatz dar, einen philosophischen Text zu verstehen. Hier geht es darum, die Schreibabsicht der Autor\*innen nachzuzeichnen (Rohbeck in Meyer 2016, 244). Dies enthält zwei Aspekte, die miteinander zusammenhängen: Erstens wird untersucht, was aus dem Gesagten oder Geschriebenen her geschlossen werden kann. Im Gegensatz zur *de dicto* Exegese wird hier also auch externe Kritik zum Text in Betracht gezogen. Zweitens wird damit ebenfalls untersucht, was die Autor\*innen unabhängig vom Text intendiert haben könnten (Brandom 2002, 100). Dies erlaubt es beispielsweise, Begriffe von Wittgenstein zu übersetzen, oder neue Begriffe einzuführen, um damit das Verständnis zu verbessern. Ebenso ist es möglich, Konklusionen zu ziehen, die über das Projekt des Tractatus hinausgehen.

Die letzte Perspektive die hier vorgestellt werden soll, kann die *traditione* Analyse genannt werden. Möchte man einen Text *de traditione* auslegen, so versucht man, den Bezug zu den Einflüssen und zum philosophischen Kanon herzustellen (Brandom 2002, 107). Es ist nicht mehr das Ziel, einzelne Begriffe oder Argumente innerhalb eines Werkes zu verstehen. Der Anspruch der *de traditione* Exegese besteht darin, die Denkschritte und Entwicklungen der Autor\*innen und des philosophischen Kanons zu erfassen (Patzig in Meyer 2016, 190). Wenn man dies an einem Beispiel angeben will, so könnte man den Einfluss Schopenhauers oder Freges auf Wittgenstein untersuchen, oder Wittgensteins Einfluss auf Anscombe oder Carnap. Des Weiteren könnte man beispielsweise auch den Tractatus in Bezug zu Wittgensteins Spätwerk setzen.

Unabhängig davon, für welche Methode man sich entscheiden wird, muss zuerst besprochen werden, wie der Gehalt eines Textes erschlossen wird. Liegt er allein im Text und wartet darauf, entschlüsselt zu werden, oder

<sup>4</sup> Wittgenstein 1921, 7.

richtet sich die Bedeutung nach den Intentionen der Leser\*innenschaft? In der vorliegenden Untersuchung wird angenommen, dass der Gehalt eines Textes dazwischen liegt (Brandom 2002, 109). Eine hermeneutische Auslegung ist so weder sture Scholastik noch purer Expressionismus. Vielmehr wird mit diesem Ansatz die Urteilskraft geschult, indem sich die Leser\*innenschaft anhand eigener Intentionen durch den Text hindurch navigieren muss (Patzig in Meyer 2016, 181). Dies geschieht in einem hermeneutischen Zirkel, wo man ständig vom Text zu den eigenen Intentionen und wieder zurück wechseln muss (Acklin 2018, 50f). Der Gehalt liegt daher im Dialog zwischen dem Text und der Leser\*innenschaft.<sup>5</sup> So entsteht offensichtlich ein Sinnpluralismus. Es können also mehrere Verständnisse eines Textes hervorgebracht werden. Dies geschieht jedoch nicht willkürlich, sondern es werden Gründe hervorgebracht, weshalb es sinnvoll ist, diese oder jene Auslegung zu akzeptieren (Brandom 2002, 105). Jene vorgebrachten Gründe stehen anschliessend wieder zur Beurteilung frei und können angenommen oder verworfen werden. Dies gilt ebenso für die Methode, die nun vorgestellt werden soll.

Die Methode welche in der vorliegenden Untersuchung verwendet wird, soll hier *metaphorischer Inferentialismus* genannt werden. Hierbei wird davon ausgegangen, dass eine einzelne, der oben dargestellten Perspektiven, nicht ausreicht, um fruchtbare Erkenntnisse aus einem Text zu gewinnen (vgl. Steinfath in Meyer 2016, 106f). Der metaphorische Inferentialismus strebt daher danach, *de dicto*, *de re* und *de traditione* Analysen abwechselnd und verbindend zu verwenden. Dies ermöglicht es, Aspekte herausheben zu können, die den einzelnen Perspektiven verborgen bleiben. In diesem Unterfangen bedarf es verständlicherweise einer gewissen Virtuosität, die gewiss auch fehleranfällig ist. Jedoch werden in der Analyse, wie bereits erwähnt, Gründe angegeben, die entweder akzeptiert oder abgelehnt werden können. Damit steht die Analyse frei zur Kritik und kann so in einen rationalen Diskurs eingegliedert werden. Des Weiteren bezeichnet der metaphorische Inferentialismus als Methode, dass man etwas *durch* etwas und etwas *als* etwas verstehen kann (Lakoff & Johnson 2018, 139). So wird bei der Interpretation die Deutungsstruktur von Phänomen (x) auf Phänomen (y) übertragen. Dadurch eröffnet sich die Möglichkeit, Phänomen (y) besser verstehen zu können, da (y) durch und als (x) verstanden wird. Dabei fallen (x) und (y) jedoch nicht zusammen, sondern Phänomen (x) wird als herme-

neutisches Instrument angewendet, um Erkenntnisse aus (y) zu gewinnen. Die Methode wird *inferentiell* genannt, weil (x) hier eine inferentielle Rolle einnimmt und in (y) eingespeist wird. So wird untersucht, welche Erkenntnisse gewonnen werden können, wenn man (y) durch und als (x) versteht (Brandom, 2002: 115). Hierbei können verschiedene Perspektiven und Kombinationen ausprobiert werden. Kreativität und Strenge müssen sich in diesem Unterfangen jedoch gegenseitig die Waage halten (Bittner in Meyer 2016, 137).

Genau dies wird in der vorliegenden Untersuchung versucht. Um Wittgensteins Verhältnis der *Welt* und der *logischen Form der Wirklichkeit* besser verstehen zu können, wird die buddhistische Realitätsvorstellung als Instrument zum Verständnis herangezogen. Das bedeutet nicht, dass Wittgenstein ein Buddhist ist, oder dass beide Auffassungen in jeglicher Hinsicht kongruent sind. Es ist jedoch möglich, Einsichten freizulegen, indem man die Deutungsstruktur der buddhistischen Realitätsvorstellung auf die Begriffe *Welt* und *logische Form der Wirklichkeit* des Tractatus überträgt. Selbstverständlich ist die Methode des metaphorischen Inferentialismus wie jede andere hermeneutische Methode kritisierbar. Es würde jedoch zu weit von der Thematik der Arbeitshypothese abweichen, dies noch ausführlicher zu besprechen. Im Wissen darum, dass also wesentliche Punkte nicht beachtet werden können, konnte die methodische Grundlage der Arbeit hier expliziert werden. Von da aus ist nun die Basis für die anstehende Untersuchung gelegt und es wird versucht, die Parallelen zwischen Wittgensteins Tractatus und der buddhistischen Realitätsvorstellung darzulegen.

## 2. Sagen und zeigen

„Ein Mönch fragte Yuanwu Keqin: „Was ist Buddha?“  
Yuanwu erwiederte: „Der Mund ist das Tor des Unglücks.““<sup>6</sup>

Die erste Parallele die aufgezeigt werden soll, kann hier *epistemologische* Ähnlichkeit genannt werden. Das bedeutet, es geht um Wahrheit und wie (und ob) diese semantisch ausgedrückt werden kann. Es ist hier nicht möglich, eine ausführliche und vollständige Darstellung der Diskurse wiederzugeben, die sich mit den Wahrheitskonzeptionen bei Wittgenstein und den unterschiedlichen buddhistischen Strömungen beschäftigen.<sup>7</sup> Es werden hier nur einzelne Aspekte aufgezeigt, welche für die vorliegende Untersuchung relevant sind.

<sup>6</sup> In Ôryû & Fukuzan 2016, 55.

<sup>7</sup> Vgl. jedoch Glock 2006, Kenny 1974, Priest 2018 und Vossenkuhl 2001.

Zuerst werden einige Erläuterungen zur Zweiteilung von *konventionaler Realität* und *ultimative Realität* im buddhistischen Sinne hervorgebracht. Sie werden in diesem Kapitel auf epistemologische Aspekte hin beleuchtet. Anschliessend folgen Bemerkungen zu Wittgenstein, wodurch aufgezeigt werden soll, dass sich hier deutliche Parallelen feststellen lassen.

Es gibt, laut einigen Philosoph\*innen, zwei verschiedene Arten von Wahrheiten: Die eine Art der Wahrheit ist verbal (semantisch), die andere ist exemplifizierend (logisch-ontologisch) und zeigt sich nur (Read in D'Amato et al. 2009, 22). So verstanden, ist eine Art der Wahrheit durch Aussagen ausdrückbar und die andere nicht. Das bedeutet dementsprechend auch, dass es Informationen oder Sinngehalte gibt, die entweder verbal oder nonverbal sind und Letztere die verbalen Wahrheiten tragen/ermöglichen/exemplifizieren (Mortensen in D'Amato et al. 2009, 6). Viele buddhistische Strömungen, darunter besonders die *Madhyamaka* und die Zen Schule, übernehmen diese Zweiteilung auch für ihre Realitätsvorstellung. Die konventionale Realität ist jene, über die man wahre Aussagen treffen kann. Hingegen ist die ultimative Realität nicht durch Worte ausdrückbar, sie zeigt sich nur. Dieser Unterschied soll hier nun kurz erläutert werden.

Die konventionale Realität ist die Welt der Myriaden von Dingen. Sie ist die vertraute und gewohnte *Lebenswelt*<sup>8</sup> des Menschen, wie sie ihm erscheint (Priest 2018, 60). Das bedeutet, man kann sie sprachlich erfassen und Aussagen über sie treffen. Durch die Sprache werden Objekte identifiziert, Attribute zugeschrieben, Subjekte den Aussagen zugeordnet und andere Unterscheidungen getroffen. Die Aussagen beziehen sich hierbei auf *Sachverhalte*, wie die Myriaden von Dingen in der Welt angeordnet sind. Diese getroffenen Unterscheidungen in Subjekt-Objekt, Objekt-Attribut etc. kommen daher zustande, weil das Äussern von Aussagesätzen stets motiviert ist, verbale Informationen zu vermitteln, welche zuerst erfasst und isoliert werden müssen (Suzuki 2019, 100). Wahrheit besteht in dieser Hinsicht, wenn die Sachverhalte mit den Aussagen auf eine Art<sup>9</sup> übereinstimmen. Hierbei können Informationen jedoch durch diese getroffenen Unterscheidungen verfälscht oder verdeckt werden (Lakoff & Johnson 2018, 18). Dies führt, so behaupten viele buddhistische Denker\*innen, zu Leid (*duhkha*) und Unglück, was in Kapitel (5) noch

genauer besprochen wird. Obwohl die ultimative Realität die konventionale Realität manifestiert, ihr also zugrunde liegt, kann die ultimative Realität nicht verbal beschrieben werden (Priest 2018, 111).

Die Wahrheit der ultimativen Realität ist nämlich keine verbale Wahrheit sondern eine exemplifizierende Wahrheit und auch die Informationen oder Sinngehalte der ultimativen Realität sind nonverbal. Fragen und Aussagen, welche die ultimative Realität betreffen sind dementsprechend fehlleitend. So ist die Frage, was Buddha sei, keine Frage, die durch die Sprache beantwortet werden kann und entpuppt sich somit als unsinnig. Die Wahrheit der ultimativen Realität ist die Exemplifizierung selbst und ist daher logisch-ontologisch (Priest, 2018: 69). Ultimative Realität strukturiert die konventionale Realität. Aussagen über die Strukturierung oder Manifestation sind etwas, das sich den Begriffen und Kategorien der konventionalen Realität entziehen. Die Wahrheit der ultimativen Realität kann daher nicht verbal erschlossen werden, sondern nur in einer Art von Bewusstsein. Dabei ist dieses Bewusstsein kein subjektbehaftetes Bewusstsein (*anatman*), sondern eine schlichte, konzeptlose Achtsamkeit, ein Wahrnehmen des Soseins (*tathata*). Dies erklärt beispielsweise auch das berühmte *Blumensutra* (vgl. Nishijima-Roshi in Dogen 2013, 8). Buddha zeigt seinen Schülern eine Blume (als Symbol des Soseins) und nur sein Schüler Kasyapa beginnt daraufhin zu lächeln. Hierauf meint Buddha, dass Kasyapa derjenige Schüler sei, der die Wahrheit der ultimativen Realität verstanden habe. Sosein kann nur gezeigt werden, man kann nichts darüber aussagen. Selbstverständlich kann man einwenden, dass hiermit Aussagen über die ultimative Realität geäussert werden können, auch wenn sie nur ex negativo sind. Dieses Paradox muss also überwunden werden. Hierauf haben unterschiedliche buddhistische Denker\*innen geantwortet. Jizang versucht beispielsweise ein System zu entwickeln, welches diesem Paradox Rechnung trägt (vgl. Priest, 2018). Die Antwort von Dogen wird im Kapitel (6) genauer untersucht. Man kann nun aufzeigen, dass Wittgensteins Begriffe *Welt* und *logische Form der Wirklichkeit* eine ähnliche epistemologische Deutungsstruktur aufweisen.

Für Wittgenstein hat das Zeigen eine zentrale Bedeutung im Tractatus (Rentsch 2003, 16). Hierbei unterscheidet er das Zeigen als Handlung in der Sprache und das Sich-zeigen als logisch-ontologische Form. Dadurch will er mit seinem Projekt die Grenze der Sprache abstecken (Wittgenstein 1921, 7). Er beschreibt, was sich überhaupt sagen lässt und was sich nur zeigt.

Mit Sätzen kann man die Welt abbilden. So ist es mög-

8 Ein Begriff der hier von Husserl übernommen wird und die hier gefassten Vorstellungen passend beschreiben kann. Vgl. hierzu Husserl 2002 und als Kommentar Precht 2012.

9 Je nach Philosophie kann diese Art der Übereinstimmung anders konzipiert und gerechtfertigt sein.

lich, dass man die Welt mit einer Menge ( $n$ ) an Sätzen beschreiben kann (TLP, 5.526). Wenn man über die Welt spricht, so verfügt man über verbale Informationen und diese Informationen können sich entweder als wahr oder als falsch herausstellen (TLP, 4.1). Dies hängt davon ab, ob der vorliegende Sachverhalt exemplifiziert ist oder nicht. Man kann hier davon ausgehen, dass Wittgenstein eine Art von Abbild- oder Korrespondenztheorie der Wahrheit im Sinn hatte.<sup>10</sup> Der Sachverhalt ( $x$ ) (nonverbal) entspricht in einer bestimmten Weise der Proposition ( $P_x$ ), die diesen Sachverhalt abbildet. Der Aussage und dem Sachverhalt ist nämlich die logische Struktur gemeinsam. Die vorigen Überlegungen zur konventionalen Realität können auch auf Wittgensteins Weltbegriff angewendet werden. Mit Sätzen sagt man etwas über die Sachverhalte der Welt der Myriaden von Dingen aus. Diese sind verbal fassbar. Sätze weisen aber nicht nur auf das, was sie aussagen (TLP, 4.461), sondern sie zeigen auch die Struktur der Welt auf. Diese Struktur der Welt ist die *logische Form*. Diese *logische Form* ist, wie der Begriff schon anmutet, nicht willkürlich, sondern strukturiert die Wirklichkeit, wie sie ist. So nimmt die logische Form der Wirklichkeit gegenüber der Welt sprachlich einen logisch-ontologischen Vorrang ein. Das bedeutet, dass die Abbildungsrelation der Aussagen zur Welt sich nur zeigen kann (Rentsch 2003, 293). Was nur gezeigt werden kann, entzieht sich jeder sinnvollen Aussage. Man kann daher die logische Form der Sätze nicht beschreiben, wie man Sachverhalte in der Welt beschreibt. Sätze, die aber etwas über die logische Form der Wirklichkeit und über die Form der logischen Elementarsätze ausdrücken, sind nicht nur selbstreflexiv und damit hinfällig (TLP, 3.332), sondern überschreiten auch die Unterscheidung von Satz und Symbol, welche Wittgenstein hervorhebt (vgl. TLP, 4.126). Sie werden dadurch unsinnig, was in Kapitel (5) besprochen wird.

Die logische Form der Wirklichkeit zeigt sich im Tractatus nicht in der Form aller sinnvollen Sätzen, sondern auch in der Struktur von Phänomenen wie beispielsweise jenes des Solipsismus oder jenes der Ethik oder der Ästhetik. Wittgenstein zieht der Sprache zwar eine Grenze, was aber nicht bedeutet, dass außerhalb der Sprache sich nichts zeigen kann. Hier wird die Parallele zur ultimativen Realität augenfällig. Die logische Form zeigt sich im Sosein ihrer logischen Struktur und umfasst damit jeden Aspekt (verbal und nonverbal) der Wirklichkeit. Man könne die logische Form der Wirklichkeit nicht anders denken, man könne nicht au-

sserhalb von ihr denken und man könne sie, so Wittgenstein, auch nicht verbal sinnvoll denken. Hinsehen allein ist das Verstehen des sich Zeigenden. So ändert Wittgensteins Philosophie auch nicht die gewöhnliche Sprache, wie es oft vermutet wird. Er will sie so belassen, wie sie ist (Read in D'Amato et al. 2009, 17). Sein Projekt im Tractatus kann daher wohl eher in Kierkegaardscher Ironie verstanden werden, das zeigt, dass es keine philosophischen Probleme gibt, die gelöst werden müssen. Philosophie in Wittgensteins Verständnis beschäftigt sich mit der logischen Form und darüber kann nichts Sinnvolles ausgesagt werden. Wittgensteins Philosophie besteht daher schlicht darin, die logische Form der Wirklichkeit aufzuzeigen (Rentsch 2003, 302). Er hält der Leser\*innenschaft nur eine symbolische Blume hin. Der Rest zeigt sich von selbst.

Dies erklärt auch, weshalb Wittgenstein am Ende seine eigenen Sätze als unsinnig bezeichnet (TLP, 6.54). Sie beschreiben die logische Form der Wirklichkeit, etwas, das sich eigentlich nur zeigen lässt. Das bedeutet nicht, dass man keine Botschaft aus seinen Sätzen erschliessen könnte. Wittgenstein strebt mit dem Tractatus vielmehr eine therapeutische Absicht an (Read in D'Amato et al. 2009, 14).<sup>11</sup> Dies wird jedoch erst in Kapitel (5) genauer besprochen. Vorerst sollen die ontologischen Parallelen zwischen der buddhistischen Realitätsvorstellung und Wittgensteins Begriffen hervorgehoben werden.

### 3. Relational und intrinsisch

*„Der Mond wird nicht nass und das Wasser wird nicht bewegt. Obgleich das Mondlicht gross und weit scheint, spiegelt es sich auf einer winzigen Fläche Wasser. Der ganze Mond und der ganze Himmel spiegeln sich in einem einzigen Tautropfen auf einem Grashalm und in einem einzigen Wassertropfen.“<sup>12</sup>*

Man könnte nun einwenden, dass mit der *epistemologischen* Ähnlichkeit noch keine bedeutenden Parallelen aufgezeigt wurden. Auch bei Heidegger (1927) findet man beispielsweise philosophische Ähnlichkeiten zum Zeigen, Sagen und Schweigen (vgl. Rentsch 2003). Es soll nun aber aufgezeigt werden, dass sich die Parallelen zwischen Wittgensteins Begriffen und der buddhistischen Realitätsvorstellung auch auf die *ontologische* Ebene erstrecken. Unter *ontologischen Parallelen* werden

<sup>11</sup> Ähnliches sehen wir auch im Rinzai Zen, wenn es um die Praxis der Koans geht (bspw. Hori in Loori 2006, 121ff). Koans selbst sind keine Rätsel, oder simple Sprachspiele. Vielmehr haben sie einen religiösen, therapeutischen Ansatz, sich mit der ultimativen Realität auseinanderzusetzen.

<sup>12</sup> Dogen 2013, 60.

10 Vgl. Kenny 1974, Rentsch 2003, sowie Sullivan in Vossenkuhl 2001.

hier zwei sich ergänzende Aspekte verstanden: Erstens geht es um die Frage, ob die menschliche Perspektive einen Einfluss auf die Ontologie der Realität hat oder nicht. Zweitens geht es auch um mereologische Fragen, also ob es metaphysische Atome (*dharma*s), kleinere und grössere Teile gibt und anschliessend ob diese Substanz respektive Buddha-Natur (*svabhava*) haben.<sup>13</sup>

Bevor nun die buddhistische Realitätsvorstellung und Wittgensteins Tractatus auf ihre ontologischen Parallelen hin besprochen werden, muss man erwähnen, dass sich die buddhistischen Strömungen in ihren metaphysischen Auffassungen und ontologischen Strukturen deutlich voneinander unterscheiden. Es werden daher nur einige Positionen aufgezeigt, die sich hier mit Wittgenstein vergleichen lassen. Dazu kommt der erschwerende Umstand, dass sich die Ontologie im Tractatus als sehr enigmatischer Teil herausstellt und die Analyse hierzu im philosophischen Diskurs weiterhin undurchsichtig geblieben ist. Dies liegt daran, dass Wittgenstein viele Begriffe nicht definiert und auch keine Beispiele anbringt, womit sie intensional und extensional unterbestimmt bleiben. Es wird hier dennoch versucht, einige Ähnlichkeiten aufzuzeigen.

Wie im vorigen Kapitel (3) schon gezeigt wurde, wird mit der konventionalen Realität im Buddhismus die Welt der Dinge verstanden und man meint daher auch die Welt, die sprachlich abgebildet werden kann. Eine wahre Aussage, die man über die Welt äussert, orientiert sich an dieser Welt, da der Satz den bestehenden Sachverhalt der Wirklichkeit abbildet (TLP, 4.023). Die konventionalen Realität ist daher stets an mindestens eine verbale Perspektive gebunden. Der Mensch ist aber allgemein an der konventionalen Realität interessiert (verbal und nonverbal) und haftet (*trsna*) ihr deshalb an (Mortensen in D'Amato et al. 2009, 9). So glaubt der Mensch auch an die Existenz eines Egos, welches eine zusätzliche Perspektive zur Welt bildet. Für die vorliegende Arbeit ist jedoch hauptsächlich das sprachliche, konzeptuelle Anhaften von Interesse. Wie schon erwähnt wurde, führen konzeptuelle Gedanken und sprachliche Ausserungen zu Verzerrungen und Verfälschungen. In der konventionalen Realität, so heisst es

in einigen Schriften im Zen-Buddhismus, erfahre der Mensch daher beispielsweise die Dinge auch als Dinge, die sich durch Substanzen auszeichnen. Substanz wird hier als ein unteilbarer, essentialer Träger der Identität von Dingen und Menschen verstanden, welche über die Zeit hinweg beständig bleibt (Han 2017, 43f). Der Mond ist daher der Mond (Trabant der Erde) und nichts anderes und dies wird er auch bleiben. So jedenfalls erscheint die ontologische Struktur der konventionalen Realität.

Die ultimative Realität stellt hingegen eine andere Realität dar. Sie wird durch eine konzeptlose Achtsamkeit (*nirvikalpa-jnana*) erschlossen (vgl. Kapitel 3). In ihr gibt es weder das Ego (*anatman*) noch konzeptuelle Unterscheidungen (Priest 2018, 57). Vielmehr ist sie wie das Wasser, das den Mond spiegelt. Sie belässt den Mond und der Mond belässt das Wasser, obwohl sie miteinander in Verbindung stehen. Der Mond ist daher auch ein Teil des Wassers, wie weiter unten gleich gezeigt wird. Die ultimative Realität ist daher ein Sosein, frei von jeder Perspektive, sei sie sprachlich konzeptuell oder an ein Subjekt gebunden (Mortensen in D'Amato et al. 2009, 10f). Dies darf nicht mit einem empirischen Objektivismus verwechselt werden. Die konzeptlose Achtsamkeit ist ein richtiges Hinsehen. Diese Perspektive, die eigentlich keine Perspektive ist, wird beispielsweise durch das Ausüben der zen-buddhistischen Praxis angestrebt (Suzuki 2019, 135). Hierzu folgt noch mehr in Kapitel (5). In den Schulen des Madhyamaka und des Zen Buddhismus wird qua richtiges Sehen in der ultimativen Realität keine zugrundeliegende Substanz der Dinge sondern die *Leere* (*ku*) erkannt (Priest 2018, 111). *Leere* ist hier nicht mit Negativität oder Absenz zu verwechseln. Vielmehr überbrückt die Leere alle Gegensätze und stellt innerhalb der ultimativen Realität einen universalen Bezug dar (Han 2017, 47). Alles ist mit allem verbunden. Nichts ist beständig. So wohnt der Mond auch im Tropfen auf dem Grashalm.<sup>14</sup>

In Wittgensteins Tractatus findet man ebenso eine solche Zweiteilung der beiden Aspekte *Perspektivität* und *Zusammenhang*. Als Erstes: Welt und Sprache bedingen sich gegenseitig (Rentsch 2003, 295). Was sich denken und sagen lässt, bildet die Grenze der Welt. Die Sprache bildet eine Grenze der Welt, aber auch das Subjekt bildet die Grenze der Welt (TLP, 5.62 und 5.632). Das Subjekt hat zwar keine Substanz, aber es zeigt sich, dass

<sup>13</sup> Eine weitere Frage die sich für die Ontologie stellen würde, wäre die Frage zur Kausalität. Diese kann hier jedoch nicht ausführlich besprochen werden. Viele buddhistische Strömungen gehen zwar davon aus, dass Kausalität zum kosmischen Prinzip des *karmas* gehört. Im Zen wird dies in der ultimativen Realität jedoch oft in Frage gestellt (vgl. hierzu Han 2017, 44, sowie Suzuki 2019, 120f). Dies wird jedoch anhand des *catuskoti* aufgelöst, da beide Aussagen zugleich wahr sein können (vgl. hierzu Priest 2018 und Kapitel 6). Auch Wittgenstein befasst sich mit der Kausalität. Er will sie zwar der Welt zuschreiben, sie jedoch nicht als logische Regel oder als Element der logischen Struktur der Wirklichkeit anerkennen (TLP, 6.35).

<sup>14</sup> Im Japanischen wird nicht der Begriff *spiegeln* verwendet, sondern Dogen sagt eigentlich *ein sich aufhalten*, ein *darin wohnen*. Hierbei wird der fundamentale Zusammenhang von Mond und Wasser deutlicher dargestellt und aus einer physikalischen Perspektive entthoben. Vgl. hierzu ausführlicher Nishijima-Roshi in Dogen 2013.

*Ich* und Welt zusammenfallen (Bell in Vossenkuhl 2001, 280). Das bedeutet nicht, dass die Welt nur aus dem Ich als Subjekt besteht, sondern Wittgenstein meint, dass das Subjekt in der Welt aufgehe, ohne ein Element davon zu werden. Die Welt ist daher die jeweils eigene Welt (TLP, 5.63). Hierin zeige sich, so Wittgenstein, die Wahrheit des Solipsismus (TLP, 5.62). Sprache und Subjekt bilden also Perspektiven, die sich auf die Welt richten. Und zweitens: Die Welt zeigt sich dem Subjekt auf ihre eigene Weise (TLP, 6.43 und Schulte in Vossenkuhl 2001, 306). Wie zeigt sich aber, so verstanden, die Welt? Sie zeigt sich im Bestehen von unabhängigen Sachverhalten.<sup>15</sup> Sie sind unabhängig, weil von einem Sachverhalt nicht auf einen anderen geschlossen werden kann (TLP, 2.061f). Die Sachverhalte sind wiederum Konfigurationen von Gegenständen, die selbst unteilbar sind (TLP, 2.02). Die Gegenstände bilden die feste, bleibende Substanz der Welt (TLP, 2.021 und 2.0271). Und weiter:

„Hätte die Welt keine Substanz, so würde, ob ein Satz Sinn hat, davon abhängen, ob ein anderer Satz wahr ist.“ (TLP, 2.022).

Auch wenn für die Gegenstände keine Beispiele angegeben werden, so müssen die Menschen etwas über ihren *Bedeutungszusammenhang*<sup>16</sup> wissen (TLP, 2.01231). Sie müssen wissen, wie die Gegenstände in Sachverhalten vorkommen können. Hier wird die Parallele zur konventionalen Realität offensichtlich. Sachverhalte und Gegenstände werden konzeptuell und pragmatisch in ihrem Bedeutungszusammenhang erfasst (Rentsch 2003, 297): Konzeptuell deshalb, weil die Sachverhalte und Gegenstände identifizierbar und kategorisierbar sind. Pragmatisch deshalb, weil die Menschen an ihrer Welt interessiert sind und ihr anhaften (*trsna*).

Anders verhält es sich mit der logischen Form der Wirklichkeit. Für Wittgenstein gibt es mehr, als das, was der Mensch in der Welt verbal zu beschreiben vermag (Rentsch 2003, 301). Dazu zählen beispielsweise Gott (TLP, 6.432), das Mystische (TLP, 6.45) und eben auch die logische Form der Wirklichkeit (TLP, 6.13). Ob dies alles dasselbe ist, muss hier nicht beantwortet werden. Die logische Form der Wirklichkeit ist frei von jeder Perspek-

tive, denn sie ist jeder Erfahrung vorgängig (TLP, 5.552). Weder das Subjekt noch die Sprache haben Einfluss auf ihre Logik. Ebenfalls ist sie von den Sachverhalten der Welt und deren Gegenständen unabhängig in dem Sinn, dass Letztere keinen kausalen Einfluss auf die Logik ausüben. Die logische Form der Wirklichkeit manifestiert sich jedoch durch die Welt. Diese Art der Manifestation ist die Leere - bei Wittgenstein wird sie *Möglichkeit* genannt. Möglichkeit als Struktur kann verschiedene Konfigurationen zulassen und die logische Form der Wirklichkeit bildet so den logischen Raum (vgl. Kapitel 5 und Kenny 1974, 91). Im logischen Raum ist alles durch die Möglichkeit universal miteinander verbunden (Metschl in Vossenkuhl 2001, 141). Der logische Raum exemplifiziert so eine Möglichkeit der Welt. Er beschreibt sie nicht verbal, aber er zeigt sie auf. Die logische Form der Wirklichkeit ist damit das, was die Welt im Innersten zusammenhält. Es scheint hier möglich, die ultimative Realität, zumindest in der Vorstellung des Zen Buddhismus, mit jener logischen Struktur der Wirklichkeit von Wittgenstein zu vergleichen. Beide sind unabhängig von einer Perspektive und beide zeichnen sich durch eine universale Verbundenheit der ontologischen Struktur aus. Von hier aus lohnt es sich, auf die dritte Parallele hinzuweisen, um die Arbeitshypothese zu vervollständigen. Es wird sich zeigen, dass Wittgenstein nicht nur epistemologisch und ontologisch ähnliche Vorstellungen wie buddhistische Strömungen aufweist, sondern auch auf einer normativen Ebene gleichgesinnt ist.

## 4. Verdeckend und klar

„In einem Sutra heisst es: ,Wenn wir alle von Natur aus vollkommene Buddhas sind, warum sind wir dann zu unwissenden, getäuschten Lebewesen geworden?’“<sup>17</sup>

Bisher wurden anhand des metaphorischen Inferentialismus sowohl epistemologische als auch ontologische Parallelen zwischen Wittgensteins Begriffen und der buddhistischen Realitätsvorstellung aufgezeigt. Die letzte Ähnlichkeit die vorgestellt werden soll, kann *normative Ähnlichkeit* genannt werden. Es geht hierbei darum, wie mit der Sprache bezüglich der Realität, der Welt und der logischen Form der Wirklichkeit umgegangen werden soll. Zuerst werden wiederum buddhistische Gedanken vorgestellt und anschliessend wird Wittgensteins Auffassung im Tractatus erläutert. Es werden sich erneut ähnliche Ansätze feststellen lassen. Der Buddhismus ist hauptsächlich eine Praxis (Read

<sup>15</sup> Das hier Probleme auftreten wird offensichtlich (Kenny, 1974: 93, sowie Priest, 2018: 43). Womöglich liegt hier beim Begriff *Unabhängigkeit* eine Äquivokation vor (vgl. TLP, 2.0122f, 2.0233, 2.601 und 6.3751). Eine andere Interpretation wäre, dass Wittgenstein selbst solche Widersprüche zulassen will (vgl. Priest 2006, 51).

<sup>16</sup> Dieser Begriff wird hier von Heidegger (1927) übernommen, da er die Vorstellungen von Wittgenstein ziemlich gut trifft. Selbst Wittgenstein hat sich 1929 positiv dazu geäußert, wie sehr seine Meinung jener von Heidegger gleicht. Vgl. hierzu ausführlich Geier 2017 und Rentsch 2003.

<sup>17</sup> In Ōryū & Fukuzan, 2016: 132.

in D'Amato et al. 2009, 17). Man bezeichnet den Buddhismus zwar oft als Religion und diskutiert über seine Philosophie aber primär wird Buddhismus (besonders in der Zen-Schule) als eine Art der Praxis verstanden. Das Wiederholen dieser Handlungen (bspw. der Meditation, *zazen*) führt zu einer Charakterveränderung (Mortensen in D'Amato et al. 2009, 12). Die Charakterveränderung besteht darin, dass der Mensch erlernt, wieder richtig sehen zu können und dabei sein Leid (*duhkha*) auslöscht (*nirvana*) (Priest 2018, 128). Hierbei ist die Charakteränderung jedoch nicht als anzustrebendes Ziel zu verstehen, denn der Mensch würde ihr dadurch nur anhaften. Vielmehr ist das Tun, also die Handlung selbst, schon Teil der Charakterveränderung, die zum richtigen Sehen und zur Auslöschung von Leid führt (Suzuki 2019, 119). Was aber ist das richtige Sehen? Dazu bedarf es einiger Bemerkungen: Es gibt nämlich zwei Arten des unrichtigen Sehens (Mortensen in D'Amato et al. 2009, 5). Einerseits kann man falsch liegen und einen Umstand nicht richtig beurteilen. Beispielsweise wenn man glaubt, dass der Satz „ $2+2=5$ “ wahr ist, obwohl er sich bei einer mathematischen Überprüfung als falsch erweist. Andererseits kann man eine falsche, konzeptuelle Vorstellung von etwas haben und das betrachtete Phänomen dadurch verdecken. So kann man beispielsweise an die Unveränderlichkeit der Welt glauben (Suzuki 2019, 116) oder an die Existenz eines Egos, welches den Identitätspol des Bewusstseinsstroms bildet (Suzuki 2019, 113). Unveränderlichkeit und Ego sind für den Buddhismus fehlgeleitete, aber nicht falsche Aussagen. Das richtige Sehen besteht nun darin, diese fehlgeleiteten Meinungen abzuwerfen (Read in D'Amato et al. 2009, 14f).

Die konzeptuelle Sprache, wie vorher erläutert, weist auf die konventionale Realität hin und verdeckt dadurch die ultimative Realität. Sie führt daher auch zu falschem Verstehen und damit zu einer unglücklichen Existenz (*samsara*). So verstanden, ist die konventionale Realität also nicht als Absenz der ultimativen Realität zu verstehen, sondern sie verdeckt sie (Priest in D'Amato et al. 2009, 81). Das richtige Sehen ist demzufolge die konzeptlose Achtsamkeit, das blosse Hinschauen, das klare Sosein (*tathata*), das mit der Praxis den Zugang zur ultimativen Realität ermöglicht, indem es die konventionale Realität abstreift (Suzuki 2019, 147f).

Auch Wittgensteins Tractatus kann als normatives Projekt aufgefasst werden. Hierfür bedarf es jedoch zuerst einige Bemerkungen zu den Sätzen und zum logischen Raum. Für Wittgenstein gibt es drei Arten von Sätzen:

„*Sinnvoll* sind empirische Aussagen der abbildenden Sprache, die wahr oder falsch sein können und daher Behauptungscharakter haben.

*Sinnlos* sind die Sätze der Logik: Sätze, die aufgrund ihrer Form allein wahr sind, mithin Tautologien. [...]

*Unsinnig* sind >metalogische< und >metaphysische< Sätze, Sätze also, die *über* die Formen der Sprache und *über* die Form des Lebens reden.“ (Rentsch 2003, 306).

Ein sinnvoller Satz, der wahr oder falsch sein kann, bestimmt einen Ort im logischen Raum (TLP, 3.4). Dabei weist der Punkt jedoch schon auf das gesamte Bestehen und die Struktur des logischen Raumes hin (TLP, 3.42). Der logische Raum ist der Raum der gesamten Möglichkeiten (Metschl in Vossenkuhl 2001, 163). Wie im Kapitel (4) schon gezeigt wurde, bildet er die Struktur der Wirklichkeit. Sinnvolle und sinnlose Sätze verweisen auf Punkte innerhalb des logischen Raumes. Anders verhält es sich mit unsinnigen Sätzen. Unsinnige Sätze sagen zwar nichts über die Welt aus, aber sie können dennoch auf etwas zeigen (bspw. auf die logische Form der Wirklichkeit). Wird sogar dieses Zeigen verfehlt, weil die logische Struktur der Wirklichkeit dadurch verdeckt wird, so kann man nach Wittgenstein von *Scheinsätzen* sprechen. Wittgenstein gesteht zwar zu, dass seine Sätze unsinnig seien (TLP, 6.54), aber sie sind keine Scheinsätze, denn er will mit seinem Projekt eine unumstößliche Wahrheit aufzeigen (Wittgenstein 1921, 7f). Daraus folgen verschiedene Implikationen. Erstens beabsichtigt er mit dem Tractatus, die verdeckten Strukturen der Sprache und damit die Form der sinnvollen Sätze freizulegen (Rentsch 2003, 14). Dies geschieht aus dem Grund, weil der Mensch viele unsinnige Sätze verwendet und einige davon auch Scheinsätze sind (Read in D'Amato et al. 2009, 21 und bspw. TLP, 4.12721). Solche Scheinsätze führen zu Verwirrung und unnötigen Problemen. Wittgenstein will daher den Menschen die Struktur der sinnvollen Sätze aufzeigen. Das Aufzeigen der verdeckten Strukturen ist jedoch nur möglich, indem er selbst unsinnige Sätze verwendet. Die Leser\*innenschaft muss daher auch den Tractatus als unsinnig verstehen oder sie versteht ihn überhaupt nicht (Read in D'Amato et al. 2009, 15f).<sup>18</sup> Wenn man den Tractatus verstanden habe, so gäbe es eigentlich nichts mehr darüber zu sagen (TLP, 7). Man muss ihn wieder abstreifen. Erst so sieht man richtig.

<sup>18</sup> Ähnlich ist auch das selbstwiederlegende Element in der Meditation. Man darf nicht auf die Charakteränderung und auf das richtige Sehen fokussiert sein und auf diese abzielen, da sonst die Meditation sich nur als verschwendete Zeit herausstellt. Vgl. hierzu ausführlicher Dogen 2013-15, Read in D'Amato et al. 2009 und Suzuki 2019.

Zweitens ist Philosophie für Wittgenstein primär eine Praxis (TLP, 4.112).<sup>19</sup> Ähnlich wie beim Buddhismus geht es nicht um die Lehre der Philosophie, sondern um die Charakterveränderung. Der Mensch lerne das richtige Sehen, indem er das Fehlgeleitete abstreift. So verhält es sich nach Wittgenstein mit dem Denken und der Sprache. Wo Fehlgeleitetes gesagt wird, sollte man besser schweigen. Hierbei wird der Mensch jedoch keiner vollständigen Änderung unterworfen. Es ist ein Lebenswandel, bei dem alles so bleibt, wie es vorher gewesen ist (Read in D'Amato et al. 2009, 23). Dies führt auch zum nächsten Kapitel (6), welches diesen Aspekt genauer beleuchtet.

## 5. Zen: Von Bergen und Wasser

*„Before I studied Zen, mountains were mountains, and water was water. After studying Zen for some time, mountains were no longer mountains, and water was no longer water. But now, after studying Zen longer, mountains are mountains, and water is just water.“<sup>20</sup>*

Wenn man wie bisher davon ausgeht, dass es Unterschiede gibt zwischen der konventionalen Realität und der ultimativen Realität, ebenso wie zwischen Wittgensteins Begriff der *Welt* und der *logischen Form der Wirklichkeit*, so liegt die Schlussfolgerung nahe, diese als unterschiedliche Arten oder Ontologien zu verstehen. Das vorliegende Kapitel will zeigen, dass dies ein Fehlschluss ist. Hierzu bedarf es zwei Erläuterungen: Erstens wird diese Aufhebung der Dualität nicht von allen buddhistischen Strömungen vertreten. Der Fokus dieses Kapitels liegt allein auf dem Zen-Buddhismus. Zweitens muss man zum Verständnis dieser Aufhebung sowohl die klassische formale Logik aufgeben, welche an dem *tertium non datur* festhält, als auch das Identitätsprinzip von Leibniz. Dadurch wird die hier präsentierte Ansicht zwar inkonsistent, jedoch nicht inkohärent (Priest in D'Amato et al. 2009, 81).<sup>21</sup> Der Zen-Buddhismus befürwortet ein Teträlemma (*catus-koti*), das bedeutet, dass Aussagen sowie ontologische Sachverhalte wahr, falsch, beides, oder weder noch sein können (Priest in D'Amato et al. 2009, 71). Es ist daher möglich, dass Unterschiedliches auch dasselbe sein kann.

---

<sup>19</sup> Vgl. hierzu ausführlich Hampe 2014.

<sup>20</sup> Quingyuan zit. nach Priest in D'Amato et al. 2009, 71.

<sup>21</sup> Dies ist kein Problem für den Zen Buddhismus, da sich dieser auf eine parakonsistente Logik berufen kann, die nicht auf der klassischen formalen Logik aufbaut. Vgl. hierzu ausführlicher Priest 2006 und 2018.

Wird man durch die Praxis von der konventionalen Realität befreit, verbleibe man, laut Zen-Buddhismus, nicht in der ultimativen Realität, welche absolut getrennt ist von der konventionalen Realität. Vielmehr ist das richtige Sehen, die konzeptlose Achtsamkeit, wie sie in den vorangegangenen Kapiteln beschrieben wurde, eine Rückkehr zur gewöhnlichen, konventionalen Realität (Priest in D'Amato 2009, 74). Man entflieht also nicht der gewöhnlichen Welt, sondern man öffnet sich ihr erst dadurch (Suzuki 2019, 135). Die ultimative Realität ist daher dasselbe wie die konventionale Realität (Priest in D'Amato et al. 2009, 79). Den Unterschied zwischen konventionaler Realität und ultimativer Realität zu ziehen und darauf zu bestehen, wäre selbst wieder ein Anhaften, welches das richtige Sehen verschleiern würde. Daher müssen solche Dualitäten aufgegeben werden (Suzuki 2019, 136). Hierbei wird auch der Unterschied zwischen Sprache und Schweigen aufgelöst (Priest 2018, 140). So sei es für Dogen auch möglich, etwas über die ultimative Realität zu sagen (vgl. Kapitel 3). Weshalb aber all diese Unterscheidungen? Sie sind ein Zeigen, welches durch das Verstehen überwunden werden muss. Berge sind also zuerst Berge, bis der Mensch durch die Praxis erkennt, dass sie leer sind. Diese Erkenntnis ermöglicht es ihm, die Berge in ihrem Sosein als Berge wieder sehen zu können. Er kehrt nach einer Charakterveränderung ins Gewöhnliche zurück, wobei alles geblieben ist, wie es war. Geändert hat sich das Wie und nicht das Was.

Wittgensteins Anschauung unterscheidet sich zwar in den Einzelheiten aber sein Anspruch kann auf eine ähnliche Absicht herunter gebrochen werden. Welt und Wirklichkeit sind dasselbe, denn die logische Form ist beiden gleich (Metschl in Vossenkuhl 2001, 160; sowie Rentsch 2003, 314ff). Die Welt und die logische Form der Wirklichkeit sind auch nicht absolut voneinander zu trennen (Rentsch 2003, 309). Denn die Logik, welche die logische Form der Wirklichkeit beschreibt, ist zweierlei: Sie ist erstens keine Lehre, sondern nur ein Hilfsmittel, um richtig sehen zu können (TLP, 6.1262). Logik ist in diesem Sinn für Wittgenstein eine phänomenologische Methode.<sup>22</sup> Sie bildet zweitens aber auch die Struktur der Wirklichkeit (vgl. Kapitel 4). Diese ist jedoch nur durch die Manifestation der Welt gegeben. Ohne Welt gibt es auch keine Wirklichkeit und umgekehrt (bikonditional/gleichursprünglich). Es gibt zwar Unsinniges und Transzendentales, dies betrifft jedoch

---

<sup>22</sup> Vgl. hierzu ausführlicher Rentsch 2003. *Phänomenologie* wird hier nicht als ein simples, wie-es-ist, verstanden sondern als die Freilegung zugrundeliegender Strukturen der Erfahrung (vgl. Prechtl, 2012).

nicht die Sprache und dazu muss sich der Mensch enthalten. Wittgenstein entfernt sich nach dieser Interpretation deutlich von der Position des logischen Positivismus, welche ihm oft unterstellt wurde.<sup>23</sup> Hat man erst einmal erkannt, dass Welt und logische Form der Wirklichkeit dasselbe sind, so erübrigen sich damit die Probleme der Philosophie. Man muss nicht mehr über sie sprechen.

Wittgensteins Tractatus besteht aus zwei Teilen, wie er selbst zugestanden hat: Das, was darin geschrieben steht und jenes, was Wittgenstein im Tractatus nicht besprechen wollte (vgl. Eilenberger 2019, 55f). Er erklärt der Leser\*innenschaft nur im Vorwort, dass die Philosophie dank ihm zu einem Ende gelangt sei (Wittgenstein 1921, 8). Der Sinn des Lebens, das Mystische, Gott, Ethik und Ästhetik sind alles Aspekte des Lebens, die weder durch die Naturwissenschaften, die alle sinnvollen Sätze bilden, beschreibbar sind noch durch die Philosophie (TLP, 6.52f). Denn die Philosophie ist nur ein Hinschauen, ein richtiges Sehen, das dem Menschen hilft, wieder zurück zum Gewöhnlichen zu finden. Wenn die logische Form der Wirklichkeit etwas Anderes wäre als die Welt, so müsste sein Buch zu einem Lehrbuch werden. Dem ist jedoch nicht so, weshalb man Wittgensteins Tractatus, sofern man ihn versteht, überwinden muss (TLP, 6.54).

## Konklusion

Das letzte Kapitel rundet die Arbeit mit einer kurzen Zusammenfassung und einigen abschliessenden Gedanken ab. Die vorliegende Untersuchung hat versucht, anhand des metaphorischen Inferentialismus das Verständnis von Wittgensteins Begriffen der *Welt* und der *logischen Form der Wirklichkeit* zu ermöglichen - Kapitel (1) und (2). Hierbei wurde die buddhistische Realitätsvorstellung zu Hilfe genommen. Kapitel (3) konnte aufzeigen, dass Parallelen auf der epistemologischen Ebene gefunden werden können. Im Kapitel (4) liessen sich zudem ontologische Ähnlichkeiten feststellen. Der normative Umgang und die Unterscheidung von Welt und der logischen Form der Wirklichkeit sind für Wittgenstein ähnlich aufgefasst, wie für den Zen Buddhismus die Realitätsvorstellung, was in Kapitel (5) und (6) aufgezeigt wurde. Die Arbeitshypothese kann dementsprechend bestätigt werden und es ist möglich, das fortwährende Projekt von Gudmunsen, Priest, Read etc. weiterzuführen, indem ein fortschreitendes Verständnis der Philosophie Wittgensteins so ans Licht gebracht werden kann.

Für den Abschluss dieser Arbeit sind hier jedoch noch einige Punkte zu ergänzen. Der metaphorische Inferentialismus muss sich notwendigerweise mit einer Abwägung von Detailanalyse und allgemeinem Überblick befassen. Die vorliegende Arbeit fokussiert vermehrt auf den Überblick, da man bei einer solchen Thematik den Wald vor lauter Bäumen nicht aus dem Blick verlieren darf. Dies ist für den Tractatus und die buddhistischen Strömungen besonders von Belang. Es war hier die Absicht, die Parallelen auf verschiedenen Ebenen aufzuzeigen (vgl. Kapitel (1)). Es wäre jedoch möglich, die hier aufgedeckten Erkenntnisse tiefer und exakter zu analysieren. Aus Platzgründen war dies nicht möglich. Die daraus folgenden Ergebnisse anschliessender Untersuchungen würden die vorliegende Untersuchung jedoch untermauern können.

Des Weiteren muss man sich stets vergegenwärtigen, dass buddhistische Strömungen und ganz besonders der Zen Buddhismus nicht durch ihre Philosophie allein verstanden werden können (Suzuki 2019, 141). Auch wenn man die Deutungsstruktur des Zen Buddhismus übernimmt, um ihn im metaphorischen Inferentialismus zu verwenden, so darf man dennoch nicht übersehen, dass der Buddhismus (und besonders der Zen Buddhismus) mehr als seine Philosophie ist und der Fokus hauptsächlich auf der Praxis liegt.

Zum Schluss sind noch Erläuterungen dazu angebracht, dass Wittgenstein selbst darauf verweist, dass bei einem richtigen Verständnis sein Werk überwunden werden müsse, wie auch im Zen Buddhismus der Anspruch erhoben wird, das Gelehrte wieder zu vergessen (Suzuki 2019, 148). An Einsichten anzuhaf-ten ist für viele buddhistische Strömungen wieder ein Verfallen, welches das richtige Sehen verumöglicht. Ein Offensein für die Welt kann sich nicht auf ein Dogma stützen und bedarf daher der Bescheidenheit, seine vermeintlich gewonnenen Erkenntnisse erneut zu hinterfragen. Im Zen Buddhismus ist das Wandern (auch symbolisch aufgefasst) ein zentraler Aspekt: Man geht über seine Erkenntnisse hinweg und haftet ihnen nicht an (Han 2017, 84). So gibt es hier noch viel zu erläutern über Wittgenstein und Buddhismus, aber darüber wird nun besser geschwiegen.

„Mich zu begleiten,  
Mein Schatten mit mir heimging  
Vom Mondbetrachten.“  
Sodô<sup>24</sup>

23 Vgl. hierzu Eilenberger 2019, Geier 2017, Read in D'Amato et al. 2009, sowie Rentsch 2003.

24 In Ulenbrook 2010, 169.

## Literatur

- Acklin, Marvin W. 2018. "Oneself as Another, Hermeneutics and Personality Assessment." *Archives of Assessment Psychology* 8 (1): 37-56.
- Brandom, Robert B. 2002. *Tales of the Mighty Dead, Historical Essays in the Metaphysics of Intentionality*. Cambridge USA: Harvard University Press.
- D'Amato, Mario, Garfield, Jay L., und Tillemans Tom J. (Hrsg.). 2009. *Pointing at the Moon, Buddhism, Logic, Analytical Philosophy*. New York USA: Oxford University Press.
- Dogen Zenji: Shobogenzo. 2013-2015. *Die Schatzkammer des wahren Dharma-Augen*. Band I-IV. 2. Auflage. Heidelberg DE: Werner Kriskeitz Verlag.
- Eilenberger, Wolfram. 2019. *Zeit der Zauberer, Das grosse Jahrzehnt der Philosophie 1919-1929*. J. G. Stuttgart DE: Cotta'sche Buchhandlung Nachfolger GmbH.
- Gadamer, Hans-Georg. 2010. *Wahrheit und Methode, Grundzüge einer philosophischen Hermeneutik*. 7. Auflage. Tübingen DE: Mohr Siebeck Verlag.
- Geier, Manfred. 2017. *Wittgenstein und Heidegger, Die letzten Philosophen*. 1. Auflage, Reinbek b. H. DE: Rowohlt Verlag GmbH.
- Glock, Hans J. 2006. "Truth in the Tractatus." *Synthese* 148 (2): 345-368.
- Gudmensen, Chris. 1986. *Wittgenstein and Buddhism*. 2. Auflage. Hounds Mills UK: Macmillan Education Ltd.
- Hampe, Michael. 2014. *Die Lehren der Philosophie, Eine Kritik*. 2. Auflage. Berlin DE: Suhrkamp Verlag.
- Han, Byung-Chul. 2017. *Philosophie des Zen-Buddhismus*. Stuttgart DE: Philipp Reclam jun. GmbH & Co. KG.
- Heidegger, Martin. 1927 (2006). *Sein und Zeit*. 19. Auflage. Tübingen DE: Max Niemeyer Verlag.
- Husserl, Edmund. 2002. *Phänomenologie der Lebenswelt, Ausgewählte Texte II*. 2. Auflage. Stuttgart DE: Philipp Reclam jun. GmbH & Co. KG.
- Kenny, Anthony. 1974. *Wittgenstein*. 1. Auflage. Frankfurt a. M. DE: Suhrkamp Taschenbuch Verlag.
- Lakoff, George, Johnson, Mark. 2018. *Leben in Metaphern, Konstruktion und Gebrauch von Sprachbildern*. 9. Auflage. Heidelberg DE: Carl-Auer-Systeme Verlag GmbH.
- Loori, John D. (Hrsg.). 2006. *Sitting with Koans, Essential Writings on Zen Koan Introspection*. Sommerville USA: Wisdom Publications.
- Meyer, Kirsten (Hrsg.). 2016. *Texte zur Didaktik der Philosophie*. 2. Auflage. Stuttgart DE: Philipp Reclam jun. GmbH & Co. KG.
- Ōryū, Genrō, Fukuzan, Imai (Hrsg.). 2016. *Koan-Sammlung, Shūmon kattōshu, Tetteki tōsui, Shōnan kattōroku*. Frankfurt a. M. DE: Angkor Verlag.
- Prechtl, Peter. 2012. *Edmund Husserl, zur Einführung*. 5. Auflage. Hamburg DE: Junius Verlag GmbH.
- Priest, Graham. 2006. *In Contradiction*. 2. Auflage. New York USA: Oxford University Press.
- . 2018. *The Fifth Corner of Four, An Essay on Buddhist Metaphysics and the Catuskoti*. Oxford UK: Oxford University Press.
- Rentsch, Thomas. 2003. *Heidegger und Wittgenstein*. Stuttgart DE: J. G. Cotta'sche Buchhandlung Nachfolger GmbH.
- Suzuki, Shunryu. 2019. *Zen-Geist, Anfänger-Geist, Unterweisungen in Zen-Meditation*. 2. Auflage. Bielefeld DE: Theseus in Kamphausen Media GmbH.
- Ulenbrook, Jan. 2010. *Haiku: Japanische Dreizeiler*. Stuttgart DE: Philipp Reclam jun. GmbH & Co. KG.
- Von Brück, Michael. 2016. *Zen, Geschichte und Praxis*. 3. Auflage. München DE: Verlag C.H. Beck oHG.
- Vossenkuhl, Wilhelm (Hrsg.). 2001. *Ludwig Wittgenstein, Tractatus logico-philosophicus*. Berlin DE: Akademie Verlag GmbH.
- Wittgenstein, Ludwig. 1921 (2018). *Tractatus-logico-philosophicus, Logisch-philosophische Abhandlung*. 37. Auflage. Frankfurt a. M. DE: Suhrkamp Verlag.

---

**Omar Ibrahim (28)** studiert Philosophie und Sozialanthropologie im Master. Er interessiert sich für Phänomenologie, Hermeneutik, Existenzialismus und dialogische Praxis.

# Der Wille zur Wahrheit im Kontext der Kritik asketischer Ideale

Untersuchung der Bedeutung eines Begriffs in  
Nietzsches *Genealogie der Moral*

## Einleitung

Durch die Geschichte der Philosophie hindurch wurden viele grundsätzliche Fragen, beispielsweise bezüglich des guten Lebens oder der Beschaffenheit der Welt, aufgeworfen und es wurden – was beinahe selbstverständlich klingen mag – Antworten auf diese Fragen gesucht. Ganz und gar nicht selbstverständlich war dieses Suchen für den Philosophen Friedrich Nietzsche. Er hat versucht Antworten auf die Frage zu finden, warum der Mensch überhaupt Antworten auf Fragen sucht und also nach Erkenntnis, Wissen, Gewissheit oder Wahrheit strebt. Gegen Ende der dritten Abhandlung seiner im Jahr 1887 erschienenen *Genealogie der Moral* hat der von ihm sogenannte „Wille zur Wahrheit“ eine zentrale Bedeutung bei seinen Erörterungen über asketische Ideale. Seiner komplexen Deutung des Willens zur Wahrheit in der *Genealogie der Moral* soll in dieser exegesischen Arbeit nachgegangen werden, indem folgende Fragestellung erörtert wird: Was ist Nietzsches Auffassung nach, der Ursprung des (unbedingten) Willens zur Wahrheit? Was ist also seine Erklärung des (unbedingten) Willens zur Wahrheit? In dieser Untersuchung wird die These vorgebracht, dass Nietzsche zufolge, die Herkunft sowie die Grundlage des (unbedingten) Willens zur Wahrheit im sogenannten „metaphysischen Glauben“ zu lokalisieren sind. Ausgehend von Nietzsches Thematisierung des metaphysischen Glaubens lassen sich seine weiteren Überlegungen zur Herkunft und Grundlage des Willens zur Wahrheit näher umreissen und nachvollziehen. Im Sinne einer Deutung der 3. Abhandlung, bezweckt die Erörterung der Fragestellung insbesondere zu erhellen, was der Zusammenhang zwischen dem Willen zur Wahrheit und dem asketischen Ideal gemäss Nietzsche ist. Zu Anfang wird gegen Larmores Auffassung des Willens zur Wahrheit als moralische Pflicht argumentiert und ein alternatives Verständnis der Beschreibung des Willens zur Wahrheit als „unbedingt“ vorgeschlagen. Daraufhin wird Nietzsches Wert-Kritik

der Wahrheit in seiner *fröhlichen Wissenschaft* in den Blick genommen und versucht aufzuzeigen, wie ihm zufolge, der Wille zur Wahrheit auf den metaphysischen Glauben zurückgeführt werden kann. Auf dieser Basis und ausgehend von der Annahme, dass das asketische Ideal eine bestimmte Wertungsweise über das Dasein verkörpert werden anschliessend grundlegende Zusammenhänge zwischen dem Willen zur Wahrheit und dem asketischen Ideal erschlossen. Zuletzt folgt eine Auseinandersetzung mit Nietzsches weiterführenden Überlegungen zur Erklärung des Willens zur Wahrheit in *Jenseits von Gut und Böse*. Dabei wird versucht, nahezulegen, dass diese Überlegungen implizit den metaphysischen Glauben zum Gegenstand haben und deshalb weiterführend sind. Anhand der Herausarbeitung dieser Überlegungen werden des Weiteren die in dieser Untersuchung behaupteten Zusammenhänge zwischen dem Wahrheitswillen und dem asketischen Ideal bestätigt. In der vorliegenden Untersuchung werden immer wieder einige Ausführungen der *Genealogie der Moral*, der *fröhlichen Wissenschaft* sowie von *Jenseits von Gut und Böse* miteinander abgeglichen und zueinander in Beziehung gesetzt.

## 1. Wille zur Wahrheit als moralische Pflicht?

Aus Nietzsches Auffassung, dass der Wille zur Wahrheit auf dem Boden der Moral gewachsen, schliesst Larmore, dass er ihm zufolge einer moralischen Pflicht gleichkomme (Lamore 2004, 165). Was Nietzsche unter Moral verstehe – woraus dann der Grund für Larmores Folgerung ersichtlich wird – sei ein „auf Mitleid (Vorrede, 6) und Ressentiment (I 10) gründendes System von allgemein verbindlichen Verpflichtungen“ (2004, 166). Daran, dass Nietzsche zufolge, der Ursprung des Willens zur Wahrheit im Moralkontext zu suchen sei, ist wenig auszusetzen. Denn, wie des Weiteren noch ersichtlich werden soll, wird dieser Gedanke von ihm explizit dar-

gelegt. Jedoch zieht Larmore daraus voreilige Schlüsse, welche den Kern von Nietzsches Gedanken zum Willen zur Wahrheit verfehlten. Der Aspekt der Verpflichtung ist in Nietzsches Auffassung über das Wesen der Moral bestimmt auch mitenthalten, aber er ist längst nicht so zentral, wie es Larmore darstellt. Es ist eine zu starke Simplifizierung, seinen vielschichtigen und komplexen Begriff der Moral mit dem Begriff eines Verpflichtungssystems gleichzusetzen. Denn Moral kann für ihn beispielsweise auch eine Form von Erkrankung bedeuten oder eine Art Maskerade sein (KSA 5, 253). Auch gilt es den historischen Horizont zu berücksichtigen, innerhalb dessen Nietzsche sowohl an älteren als auch an zeitgenössischen Moralvorstellungen Kritik übt. Der von Larmore erwähnte Begriff der Verpflichtung ist im Wesentlichen zwar auf Kant zurückzuführen, mit dessen Philosophie sich Nietzsche auseinandergesetzt hat, aber es bleibt letzten Endes offen, ob und wie eng ein kantisches Moralverständnis in einem Zusammenhang mit Nietzsches Moralauffassung steht.

Larmore scheint Moralvorstellungen der heutigen Moderne in die Rekonstruktion von Nietzsches Denken hineinzuprojizieren. Dieser Verdacht drängt sich angesichts von Larmores Aussage auf, dass bei Nietzsche der Wille zur Wahrheit – als eine Pflicht – gegenüber den Mitmenschen einzuhalten sei (Larmore 2004, 165). Bei einer solchen Interpretation ist eine gewisse Färbung durch die moderne Auffassung, dass das Erfüllen von Pflichten anderen gegenüber, zum Schutz deren Rechte auf Freiheit und Wohlergehen, geschehe, unübersehbar.<sup>1</sup> Was in erster Linie gegen eine solche Auslegung spricht, ist das Fehlen jeglichen Hinweises in der *Genealogie der Moral*, welcher auf ein Verständnis des Willens zur Wahrheit als eine Art Pflicht, welche man anderen gegenüber zu erfüllen habe, hindeuten würde. Überhaupt bringt Nietzsche den Willen zur Wahrheit nirgendwo, weder in der *Genealogie der Moral* noch in seinen früheren Werken, in einen direkten Zusammenhang mit dem Begriff der Verpflichtung.<sup>2</sup> Mit dem Augenmerk auf eine bestimmte Stelle in der Genealogie der Moral, wo Nietzsche schreibt, dass der Glaube an die Wahrheit zu Asketismus verpflichtet, wäre man vielleicht geneigt, zu widersprechen (KSA 5, 399). Es

ist aber falsch, den Glauben an die Wahrheit mit dem Willen zur Wahrheit gleichzusetzen. Denn wenig später präzisiert Nietzsche, dass der Wille zur Wahrheit – verstanden als „Glaube an das asketische Ideal“ – zum Glauben an die Wahrheit als dessen „unbewusster Imperativ“ zwinge (KSA 5, 400). Dieser zentralen Aussage gilt es in Bezug auf die Ursprungsfrage Aufmerksamkeit zu schenken, was weiter unten geschehen wird. Es ist darauf zu beharren, dass hier nicht von einem Willen zur Wahrheit als moralischer Verpflichtung, die man anderen gegenüber einzuhalten hat, die Rede ist. Stattdessen ist anzunehmen, dass der unbewusste Zwang, von dem die Rede ist, psychologischer – und damit eher deterministischer Natur ist. Dass ein Wille zu etwas zu einem Glauben an dieses selbe etwas zwinge, ist für sich genommen vielmehr als eine psychologische Behauptung aufzufassen, in Bezug auf die es keinen Anlass gibt, eine Verbindung zum Begriff der moralischen Verpflichtung herzustellen. Was aber gerechtfertigten Anlass zu einer Verbindung des Willens zur Wahrheit mit dem Begriff der moralischen Verpflichtung gibt, ist die Charakterisierung des Wahrheitswillen als „unbedingt“. Und vielleicht war es dies, was Larmore zu seiner Deutung verführt hat. In der Tat liegt die Vermutung nahe, Nietzsche meine, dass der Wille zur Wahrheit eine moralische Pflicht darstelle, da mit dem Wort „unbedingt“ auch der Geltungsanspruch moralischer Pflichten hervorgehoben wird. Doch was, wenn die Unbedingtheit im Kontext von Nietzsches Kritik des Willens zur Wahrheit eine gänzlich andere Bedeutung hat? Im nachfolgenden Abschnitt soll nun eine alternative Erklärung für diese Charakterisierung des Willens zur Wahrheit präsentiert werden, womit erste Antworten auf die Ursprungsfrage gegeben werden.

## 2. Wert und Unbedingtheit

Um dem Grund für Nietzsches Bezeichnung des Willens zur Wahrheit als „unbedingt“ auf die Spur zu kommen, ist es notwendig, auf die Stelle in der *Genealogie der Moral* einzugehen, wo er den Glauben an die Wahrheit aus dem Willen zur Wahrheit ableitet. Der Wille zur Wahrheit wird hier zum einen, als Glaube an das asketische Ideal spezifiziert, und in ergänzender Weise als Glaube an einen metaphysischen Wert, beziehungsweise an einen „An-sich“-Wert der Wahrheit (KSA 5, 400). Aber was hat man sich darunter vorzustellen? Ein vielversprechender Ansatzpunkt für die Beantwortung dieser Frage bietet sich zwei Seiten später an, wo Nietzsche die „Überschätzung der Wahrheit“ als den gemeinsamen Boden von Wissenschaft und asketischem Ideal bezeichnet (KSA 5, 402). Es ist davon auszugehen, dass

1 Kersting, W.: „Pflicht.“ In Historisches Wörterbuch der Philosophie, Bd. 7. Darmstadt 2019, Sp. 432.

2 Erst die Götzen-Dämmerung, in der die „wahre Welt“ als eine Art Verpflichtung charakterisiert wird, könnte einen solchen Zusammenhang ansatzweise nahelegen (KSA 6, 80-81). Aber auf eine Berücksichtigung späterer Werke wird bewusst gänzlich verzichtet. Eine solche würde die Dynamik in Nietzsches Denken über seine Werke hinweg ausser Acht lassen und birgt die Gefahr Ideen in die *Genealogie der Moral* hineinzuprojizieren, die eigentlich nur den späteren Werken zuzuschreiben sind.

das, was hier „Überschätzung der Wahrheit“ genannt wird, wesensgleich mit dem Glauben an den metaphysischen Wert der Wahrheit ist. Dies deshalb, weil hinsichtlich der Bedingtheit des Glaubens an die Wahrheit durch den Willen zur Wahrheit als Glaube an einen metaphysischen Wert der Schluss gezogen wird, es gebe keine voraussetzungslose Wissenschaft (KSA 5, 400). Die Wissenschaft hat demnach notwendigerweise einen Boden, auf dem sie steht und dieser Boden ist die Überschätzung der Wahrheit oder synonym formuliert: der Wille zur Wahrheit. Auch die beiläufige Bemerkung „ich gab dies schon zu verstehen“, nämlich dass die Wissenschaft und das asketische Ideal auf dem gemeinsamen Boden der Wahrheitsüberschätzung stehen, bestärkt in der Annahme einer äquivalenten Beziehung zwischen der Überschätzung der Wahrheit und dem unbedingten Willen zur Wahrheit (KSA 5, 402).

Insofern hier das Eine dem Anderen entspricht und unter Rücksichtnahme auf Nietzsches Deutung des Wahrheitswillen als einen Glauben an einen metaphysischen Wahrheitswert, kann der metaphysische Wert am ehesten als „unabschätzbarer Wert“ und „unkritisierbarer Wert“ umformuliert werden. Das ergibt sich daraus, dass die Überschätzung der Wahrheit als „Glaube an die Unabschätzbarkeit, Unkritisierbarkeit der Wahrheit“ präzisiert wird (KSA 5, 402). Der Glaube an den metaphysischen Wert der Wahrheit, ist also auch ein Glaube an die Unabschätzbarkeit und Unkritisierbarkeit der Wahrheit. Es sei darauf hingewiesen, dass eine solche Deutung der *Genealogie der Moral* derjenigen Larmores widerspricht, welcher meint, dass Nietzsche die Herkunft des Willens zur Wahrheit im asketischen Ideal sehe (Larmore 2004, 166). Da das Verhältnis zwischen dem Willen zur Wahrheit und dem asketischen Ideal in der *Genealogie der Moral* aufgrund der Komplexität der Bedeutung des asketischen Ideals letztlich nicht richtig erfassbar ist, bleibt offen, ob das Eine im Anderen seinen Ursprung hat oder umgekehrt.<sup>3</sup> Es gäbe auch Grund zur Annahme, dass es insofern gar kein Verhältnis zwischen dem Willen zur Wahrheit und dem asketischen Ideal gibt, als dass der Wille zur Wahrheit sogar mit dem asketischen Ideal identifiziert wird. Nietzsche sagt nämlich, der Wille zur Wahrheit sei eigentlich der Kern des asketischen Ideals (KSA 5, 409). „Kern“ kann dabei als

das eigentliche Wesen einer Sache verstanden werden. Wird nun der metaphysische Wert fortan als unkritisierbarer Wert verstanden, dann wird Nietzsches Tendenz erkennbar, den Wertglauben als dogmatisch zu betrachten. Die Unbedingtheit des Willens zur Wahrheit wäre damit auf die Unbedingtheit eines Dogmas zurückzuführen, das womöglich für Nietzsche in Hinblick auf den Wert der Wahrheit besteht. Vielleicht sollte aber der Fokus vorerst darauf liegen, diese Unbedingtheit in Zusammenhang mit der Unabschätzbarkeit des Wahrheitswerts, an die geglaubt werde, zu betrachten. Ein Seitenblick auf den zweiten Aphorismus des fünften Buchs der *fröhlichen Wissenschaft* – ein Aphorismus dessen Lektüre Nietzsche selbst dem Leser zum Verständnis der Ausführungen in der *Genealogie der Moral* ans Herz legt – betont diesen Zusammenhang. Die Bejahung der Wahrheit, also der Wille zur Wahrheit, sei in der modernen Wissenschaft dergestalt, dass darin die Überzeugung zum Ausdruck komme, es gebe nichts wichtigeres als Wahrheit und alles andere sei seinem Wert nach ihr gegenüber zweitrangig (KSA 3, 575). Die Wahrheit sei nicht nur erstrebenswert, sondern habe darüber hinaus den Sonderstatus eines unaufwiegbareren und damit eben unabschätzbaren Werts. „Wahrheit um jeden Preis“ ist nach Nietzsche die Überzeugung, welche für die Wissenschaft als deren grundlegende Voraussetzung gelte (KSA 3, 576). Da sich der menschliche Wille nach Wertvorstellungen ausrichtet, macht eine solche superlativische Priorisierung der Wahrheit als Wert den Willen zur Wahrheit unbedingt. Als Zwischenresultat lässt sich festhalten, dass das, was in der *fröhlichen Wissenschaft* als die Voraussetzung der Wissenschaft gesehen wird, ziemlich genau dem zu entsprechen scheint, was später in der *Genealogie der Moral* als Glaube an den metaphysischen Wert der Wahrheit bezeichnet wird. Diese Behauptung wird auch dadurch gestützt, dass Nietzsche die Möglichkeit einer Wissenschaft ohne einen Glauben, den sie zur Voraussetzung hat, abstreitet. Die Überzeugung an einen unbedingten Wert der Wahrheit, beziehungsweise eben der Glaube an den metaphysischen Wert der Wahrheit, kommt also einerseits durch den Willen zur Wahrheit zum Ausdruck und zugleich muss er auch dasjenige sein, was dem Willen zur Wahrheit zugrunde liegt.

In der *fröhlichen Wissenschaft* geht es Nietzsche darum, die herkunftstheoretische Grundlage des unbedingten Willens zur Wahrheit aufzudecken. Methodisch vollzieht er seine Untersuchung mittels einer Infragestellung der Wahrheit als Wert, und damit durch eine Infragestellung der Rechtfertigung des Willens zur Wahrheit. In der *Genealogie der Moral* wird dieser

3 Der Titel der dritten Abhandlung „Was bedeuten asketische Ideale?“ gibt bereits zu verstehen, dass es das Ziel dieser Abhandlung ist, dem Leser näherzubringen was überhaupt alles unter den Begriff des asketischen Ideals fällt (KSA 5, 339). Es wäre daher widersinnig mit einer bestimmten Definition hinsichtlich dieses Begriffs zu arbeiten. Aus diesem Grund wird die Frage nach dem Verhältnis zwischen dem Willen zur Wahrheit und dem asketischen Ideal vorerst ausser Acht gelassen. Im weiteren Verlauf wird aber über diese Relation erhellt werden.

Ansatz dann indirekt in Form einer Aufforderung formuliert (KSA 5, 401). Es könnte eingewendet werden, dass in der *Genealogie der Moral* nicht klar ersichtlich ist, ob der Appell zur Kritik des Willens zur Wahrheit die Klärung der Frage nach dessen Ursprung wirklich miteinschliesst oder ob sich die Kritik in einer normativen Auseinandersetzung mit der Wahrheit als Wert erschöpft. Als Antwort darauf soll mit Verweis auf das Vorwort, in dem geschildert wird, dass eine Wertkritik ein Verständnis der Entstehung der Werte, die im Fokus stehen, voraussetzt, zumindest der enge Zusammenhang zwischen der Wertfrage und der Ursprungsfrage unterstrichen werden (KSA 5, 253). Die Tendenz zu dieser Methodik macht sich des Weiteren auch zu Beginn von *Jenseits von Gut und Böse* bemerkbar, wobei da zwar nicht vom Hinterfragen des Werts der Wahrheit, sondern von der Frage nach dem Wert des Wahrheitswillen gesprochen wird. Dort heisst es, dass die Wertfrage die „gründlichere Frage“ sei als die Frage nach dem Ursprung des Willens zur Wahrheit (KSA 5, 15). Das macht deutlich, dass für Nietzsche eine Kritik des Werts der Wahrheit den Weg für eine Kritik über den Ursprung des Willens zur Wahrheit ebnet. In der *Genealogie der Moral* wird insbesondere auch die Notwendigkeit einer Rechtfertigung der Wissenschaft, beziehungsweise des ihr zugrunde liegenden Willens zur Wahrheit betont. So deutet er an, dass die Wissenschaft aus dem Glauben an den metaphysischen Wert der Wahrheit heraus ihr „Recht auf Dasein“ habe und gegen Ende des Aphorismus gelangt Nietzsche zur Feststellung, dass eine solche Rechtfertigung gänzlich fehle, aber dennoch nötig sei (KSA 5, 400-401). Im nächsten Abschnitt soll in Grundzügen herausgearbeitet werden, weshalb Nietzsche zufolge, diese Rechtfertigung fehlt.

### **3. Nietzsches Kritik am Wert der Wahrheit**

Um zu klären, was unter einem metaphysischen Wert in Bezug auf die Wahrheit zu verstehen ist, ist es notwendig den argumentativen Gedankengang zu rekonstruieren, der in der *fröhlichen Wissenschaft* hinsichtlich des Wahrheitswillen gemacht wird. Eine solche Rekonstruktion gewährt einen direkten Einblick in die Art und Weise, wie Nietzsche zu seiner These über den Ursprung des Willens zur Wahrheit kommt. Es sei vorweggenommen, dass Nietzsche zum Schluss kommt, der Wille zur Wahrheit habe – in der unbedingten Weise, in der er sich aussere – keine rationale Rechtfertigung und sei daher ein irrationales Phänomen. Dazu kommt er durch folgende Überlegungen: Der Wille zur Wahrheit könne entweder als passives Sich-nicht-täuschen-lassen-Wollen oder als aktives Nicht-täuschen-Wollen in-

terpretiert werden (KSA 3, 575). Als der Wille, sich nicht täuschen zu lassen wäre der Wille zur Wahrheit Skeptizismus schlechthin, der aber nur unter der Bedingung entstanden sein könne, dass es für eine kontinuierliche skeptische Einstellung zum Leben einen prudentiellen Grund geben müsse (KSA 3, 575). Es zahlt sich aber gemäss Nietzsche nicht aus in ständiger Skepsis zu leben, denn auch der eigene Irrtum könne sich im Leben als nützlich erweisen, und es bleibe für den Einzelnen ungewiss, ob gesamthaft das unbedingte Misstrauen gegenüber dem unbedingten Zutrauen einen Vorteil biete (KSA 3, 575-76). Damit wird die erste Interpretationsweise ausgeschlossen und der Wille zur Wahrheit als der Wille, nicht zu täuschen, festgelegt. Als solcher müsse der Wille zur Wahrheit seinen Ursprung in der Moral haben, denn die aktive Täuschung, welche auch die Selbstdäuschung miteinschliesst, sieht Nietzsche als eine zentrale Essenz des Lebens an, welche eine auf rationalen Gründen der Klugheit basierende Berechtigung und Herkunft dieses Willens ausschliesst (KSA 3, 576). Implizit möchte Nietzsche seiner Leserschaft insbesondere Folgendes mitteilen: Im Unterschied zu einer aus seiner Sicht möglichen Bereitschaft des menschlichen Willens, sich täuschen zu lassen, welche er als Zutraulichkeit charakterisiert und bei der unklar bleibt, ob diese sich gegenüber einer Einstellung des strengen Misstrauens auszahlt, ist der Wille, zu täuschen, gegenüber dem Willen nicht zu täuschen, ganz eindeutig vorteilhafter. Das Element der (aktiven) Täuschung anderer und sich selbst, ist für Nietzsche so stark in der Lebenswelt verankert, dass sich für ihn der Wille zur Wahrheit als ein Wille, nicht zu täuschen wie ein „lebensfeindliches zerstörerisches Prinzip“ ausnimmt (KSA 3, 576). Er spekuliert: „„Wille zur Wahrheit“ – das könnte ein versteckter Wille zum Tode sein. – “ (ebd.). Insofern der Wille zur Wahrheit, so Nietzsche, etwas Lebensfeindliches sei, bringe er die Bejahung einer jenseitigen, metaphysischen Welt, die der Lebenswelt entgegengesetzt sei, zum Ausdruck (KSA 3, 576-77). Der Wille zur Wahrheit wird als ein Wille zum Metaphysischen gedeutet. Der Zusammenhang mit der Moral als Ursprung des Willens zur Wahrheit ergibt sich nun in bedeutsamer Weise aus Nietzsches Selbstverständnis, dass das Leben seinem Wesen nach unmoralisch sei (ebd.). Damit stuft er die Moral, beziehungsweise moralische Überzeugungen, gleich dem Willen zur Wahrheit ein, erstens als lebensverneinend und zweitens als Metaphysik-bejahend. Die Rückführung des Willens zur Wahrheit auf die Moral wird gegen Ende des Aphorismus 344 dahingehend zugespitzt, dass der sogenannte Glaube an die Wissenschaft bedingt sei durch den me-

taphysischen Glauben, die Wahrheit sei göttlich (KSA 3, 576). Analog zu den bisher diskutierten Stellen in der *Genealogie der Moral* sollte man sich durch diese Behauptung auf alle Fälle nicht dazu verleiten lassen, Differenzierungen zu machen, wo keine angebracht sind. „Glaube an die Wissenschaft“ ist nur eine andere Bezeichnung für jene Überzeugung, die Nietzsche als Voraussetzung der modernen Wissenschaft sieht und die durch den Willen zur Wahrheit zum Ausdruck gebracht wird; nämlich, dass die Wahrheit wichtiger als alles andere sei (KSA 3, 576). Zumal der Wille zur Wahrheit selbst auf dieser Überzeugung ruht, hat er seinen Ursprung also im metaphysischen Glauben. Ebenso steht die Bezeichnung „Glaube an einen metaphysischen Wert“ die in der *Genealogie der Moral* verwendet wird, wie im vorigen Abschnitt schon festgehalten wurde, für die Überzeugung, dass der Wahrheit ein unverhandelbarer Wert zukommt, und sollte nicht mit dem metaphysischen Glauben verwechselt werden. Wie in der *fröhlichen Wissenschaft*, in welcher der Glaube an die Wissenschaft als Voraussetzung der Wissenschaft selbst charakterisiert wird, so trifft dieselbe Charakterisierung in der *Genealogie der Moral* in Bezug auf den Glauben an einen metaphysischen Wert zu. Dies legt nahe, dass das Eine dem Anderen entspricht.

Daraus wird nun ersichtlich, was in der *Genealogie der Moral* mit dem metaphysischen Wert der Wahrheit gemeint ist. Dem Wert der Wahrheit wird insofern ein metaphysischer Status zugestanden, als an die Wahrheit selbst, wie an Gott geglaubt wird. Die unbedingte Hochschätzung der Wahrheit wäre demnach auf eine gewisse Gottesfurcht zurückzuführen. In der *Genealogie der Moral* bringt Nietzsche – anschliessend an die Selbstzitierung des Endes von Aphorismus 344 der *fröhlichen Wissenschaft* – nochmals auf den Punkt, worin genau der Bezug zur Moral für ihn besteht: Insofern Wahrheit als göttlich oder als Gott selbst gesetzt wird, darf die Wahrheit selbst in keinerlei Hinsicht problematisiert werden (KSA 5, 401). Er hebt an dieser Stelle speziell das Verb „dürfen“ hervor und fügt sogar die Frage an, ob man dieses Wort verstehne. Damit möchte er offensichtlich die moralische Dimension des Willens zur Wahrheit dem Leser klarmachen. Diese Tabuisierung, so scheint es, möchte Nietzsche auch hervorheben, wenn er von der Unabschätzbarkeit und Unkritisierbarkeit der Wahrheit spricht. Das Anzweifeln des unbedingten Wahrheitswerts wäre christlicher Vorstellung nach ein Zweifelhegen – und damit einhergehend ein Unglaube – gegenüber Gott, welches nicht erlaubt ist. Nietzsche zeigt sich überzeugt, dass sich das Problem vom Wert der Wahrheit stelle, sobald der metaphysische Glaube

an Gott als die Wahrheit nicht mehr bestehe (KSA 5, 401). In dramatischer Weise schildert er, dass der metaphysische Glaube immer unglaublichwerde, je mehr sich Gott selbst als die „längste Lüge“ herausstelle (ebd.). Eine Rechtfertigung des Wahrheitswillens kann also weder aus dem Nutzen-Aspekt heraus noch über einen metaphysischen Weg geschehen, da Gott als metaphysisches Konstrukt zunehmend unglaublich wurde, oder anders formuliert, da „der alte Gott“ tot ist (KSA 3, 573-74). Gegen Ende von Aphorismus 344 lässt sich feststellen, dass Nietzsche den Willen zur Wahrheit als womöglich bedingt durch den Irrtum und durch das Lügenhafte zu betrachten scheint – ein Gedanke, mit dem Nietzsche verstärkt in *Jenseits von Gut und Böse* spielt und der thematisiert werden muss, um die vollständige Reichweite der Auseinandersetzung Nietzsches mit der Ursprungsfrage zu erfassen (KSA 3, 577).

#### 4. Die asketische Wertung des Daseins

Die bisherigen Erörterungen über das, was in der *Genealogie der Moral* unter dem metaphysischen Wert der Wahrheit verstanden wird, lassen nun auch Rückschlüsse darauf zu, weshalb der Wille zur Wahrheit als Glaube an den metaphysischen Wert der Wahrheit gleichzeitig den Glauben an das asketische Ideal selbst bedeutet. Eine der vielen Facetten des asketischen Ideals ist für Nietzsche, dass es eine bestimmte Wertungsweise verkörpert. Durch diese Wertungsweise, welche durch den sogenannten Typus des asketischen Priesters vermittelt wird, werde das Leben in eine gegensätzliche und logisch-ausschliessende Beziehung zu einer anderen Form des Seins gesetzt (KSA 5, 362). Eben diese Gegensatz-Beziehung ist dieselbe, welche in der *fröhlichen Wissenschaft* im Kontext der Kritik des Werts der Wahrheit bereits formuliert wurde. Die Verneinung des Lebens ist ein zentraler Aspekt des asketischen Ideals als Wertungsweise und sie äussert sich in einer asketischen Lebensweise. Auffällig ist, dass sowohl gemäss der *Genealogie der Moral* als auch nach der fröhlichen Wissenschaft die Verneinung des Lebens erst aus der Bejahung des Metaphysischen heraus resultiert (KSA 3, 577). Gemäss Nietzsche versteht die asketische Person ihre Lebensweise als praktische Widerlegung des Lebens als Irrtum, um dadurch in eine metaphysische Welt zu transzendentieren (KSA 5, 362). Von Relevanz ist also insbesondere, dass diese Gegensatz-Beziehung von der Welt des Lebens zur metaphysischen Welt in einem logischen Sinn zu verstehen ist. So ist es verständlich, weshalb die Bejahung der metaphysischen Welt als des Wahren notwendig mit der Verneinung des Lebens als des Falschen einhergeht. Diese logische Gegensätzlichkeit ist insbesondere in Bezug auf weitere

Hypothesen Nietzsches zum Ursprung des Wahrheitswillen äusserst relevant, wie im nächsten Abschnitt verdeutlicht werden sollte.

Basierend auf den bisherigen Erörterungen kann die These aufgestellt werden, dass für Nietzsche der Wille zur Wahrheit als Wertschätzung einer metaphysischen Welt nicht nur auf der asketischen Wertungsweise basiert, sondern gleichzeitig die wesentliche, lebenspraktische Ausdrucksform derselben ist. Dadurch ist dieser Wille auch essentieller Bestandteil des asketischen Ideals selbst. So betrachtet ist der Wille zur Wahrheit also Ausdruck des Glaubens an das asketische Ideal. Eine solche These kann durch zwei weitere Aspekte ergänzt werden: Zum einen behauptet Nietzsche in der *Genealogie der Moral* den Willen zur Wahrheit als Kern des asketischen Ideals und damit als dessen Verkörperung. Insofern das asketische Ideal nebst anderem als eine bestimmte Wertungsweise aufgefasst werden kann, ist diese These also naheliegend. Zum anderen lässt sich mit zahlreichen Stellen belegen, dass für Nietzsche das Erkenntnisstreben, selbst im Kontext der modernen Wissenschaftlichkeit, immer mit asketischen Charakterzügen einhergeht. Gemeint ist dabei unter anderem das von ihm sogenannte Verzichtleisten auf Interpretation, denn die Interpretation enthält Nietzsche zufolge, unwissenschaftliche Elemente, wie das „Zurechtschieben“, „Weglassen“ und „Umfälschen“. Eben dieser Verzicht auf Interpretation ist bedeutend im Zusammenhang mit der von ihm aufgestellten historischen Diagnose einer Wandlung und Sublimierung des „christlichen Gewissens“ hin zum „wissenschaftlichen Gewissen“ (KSA 5, 409). Ein weiterer Zug ist das Kühlstellen der Affekte als Voraussetzung für Wissenschaftlichkeit (KSA 5, 403). Darauf kann im Rahmen dieser Arbeit zwar nicht näher eingegangen werden, doch soll damit auf die hohe Bedeutsamkeit der praktischen Ausdrucksform und auch der Wirkung des Willens zur Wahrheit zumindest ansatzweise hingewiesen werden.

## 5. Irrtum, Täuschung und Unwissenheit

Nietzsches Gedanken zum Ursprung des Willens zur Wahrheit bleiben nicht bei der These stehen, dass es ein metaphysischer Glaube ist, auf dem der Wille zur Wahrheit beruht. Vielmehr findet sich eine vertiefte Auseinandersetzung mit dem metaphysischen Glauben im zweiten Aphorismus in *Jenseits von Gut und Böse*. Eine Beschäftigung mit diesem Aphorismus ist sehr aufschlussreich, um ein reichhaltigeres Bild von Nietzsches weiteren Überlegungen zum Ursprung des Willens zur Wahrheit zu gewinnen. Darin werden hypothetische Erwägungen zur Sprache gebracht, die nicht

als Behauptungen missverstanden werden sollten. Zu Beginn dieses Aphorismus wird ein metaphysisches Argument skizziert, welches das Urteil begründet, „die Dinge höchsten Wertes“, unter welche die Wahrheit fällt, hätten ihren Ursprung in Gott, im Ding an sich (KSA 5, 16). Das Argumentationskonstrukt als Gesamtes ist für Nietzsche Ausdruck eines abstrakten Glaubens, den er als „Glaube an die Gegensätze der Werte“ bezeichnet (ebd.). Dieser Glaube, welcher das Urteil des metaphysischen Glaubens begründet, da er dessen Gehalt als seine Schlussfolgerung enthält, beinhaltet zwei aus Nietzsches Sicht zweifelhafte Annahmen. Zum einen werde davon ausgegangen, dass es sich beim Willen zur Wahrheit, der Wahrheit und der selbstlosen Handlung um Dinge höchsten Werts handle, während einem sogenannten Willen zur Täuschung, dem Irrtum und dem Egoismus gar kein Wert zugesprochen werde (ebd.). Zum anderen werde angenommen, diese Werte verhielten sich gegensätzlich zueinander. Im Sinne einer wohlwollenden Interpretation ist diese Prämisse so zu verstehen, dass insbesondere die Träger solcher Werte im Gegensatz zu einander stehen. Als Drittes liegt, Nietzsches Darstellung gemäss, dem metaphysischen Argument die Annahme zu Grunde, dass die Lebenswelt täuschend, irrtümlich und egoistisch sei (ebd.). Während Nietzsche von der dritten Annahme stillschweigend absieht und er dieser zuzustimmen scheint, kritisiert er bezüglich der zweiten Annahme, dass es berechtigt sei, grundlegend in Frage zu stellen, ob es Gegensätzlichkeit und im Besonderen Wertgegensätze gebe (KSA 5, 16). Hinsichtlich der ersten Annahme stellt sich für Nietzsche die Frage, ob diese Art und Weise der Wertschätzung nicht bloss eine vordergründige Wertung darstelle, der alternative Wertungsperspektiven gegenübergestellt werden könnten (ebd.). Nietzsche zieht die Möglichkeit einer wertenden Perspektive in Betracht, die dem Irrtum und der Täuschung einen höheren Wert als der Wahrheit zuspricht und diesen Dingen damit eine positivere Bedeutung beilegt (KSA 5, 16-17). In der Ausführung des metaphysischen Arguments sind Grundzüge der asketischen Wertungsweise, die in der *Genealogie der Moral* beschrieben werden, deutlich wieder erkennbar. Ein erster Gesichtspunkt, der einen engen Zusammenhang zur asketischen Wertungsweise vermuten lässt, ist, dass dem metaphysischen Argument eine ganz bestimmte Wertung zugrunde liegt, die mit der asketischen Wertungsweise identifiziert werden kann, und zwar schlicht deshalb, weil diese Wertung auf analoge Weise die Lebenswelt abwertet und im Gegenzug die metaphysische Welt aufwertet (KSA 5, 362). Der Bezug wird des Weiteren darin deutlich, dass hier

wie dort die Gegensätzlichkeit jeweils in einem logisch ausschliessenden Sinne zu verstehen ist. Denn innerhalb des metaphysischen Arguments wird die Möglichkeit bestritten, dass Gegensätze auseinander entstehen können und am Ende wird die Unabteilbarkeit der höheren Werte aus der mit den gegensätzlichen Werten behafteten Lebenswelt behauptet. Insofern das von Nietzsche skizzierte Argument, welches den metaphysischen Glauben an den Ursprung der Wahrheit in Gott begründet, auf der asketischen Wertungsweise aufbaut und der metaphysische Glaube die Stütze des Willens zur Wahrheit ist, ergibt sich, dass der Wille zur Wahrheit mitunter aus der asketischen Wertung des Lebens entstanden ist.

Von weitreichender Bedeutung für die Frage nach dem Ursprung des Willens zur Wahrheit ist Nietzsches Spekulation über den abstrakten Gedanken, dass grundsätzlich das, was unter anderem den Wert der Wahrheit bestimmt, in einem direkten Zusammenhang mit dem Irrtum und der Täuschung steht oder mit diesen gar identisch ist (KSA 5, 17). Da die Frage nach dem Wert bei Nietzsche mit der Ursprungsfrage eng verknüpft ist, liegt der Gedanke nahe, dass Wahrheit und der Wille zur Wahrheit auf dem Irrtum und dem Willen zur Täuschung fussen. Obwohl Nietzsche diesbezüglich alles andere als konkret wird, spielt er zu Anfang des Aphorismus unübersehbar auf eben diese Idee an, wenn er die Frage aufwirft, wie Wahrheit aus dem Irrtum und Wille zur Wahrheit aus dem Willen zur Täuschung heraus entstehen können (KSA 5, 17). Doch worin liegt Nietzsche zufolge nun genau der Irrtum, die Täuschung und das Unwissen, auf dem der Wille zur Wahrheit womöglich gründet? Antworten auf diese Frage kristallisieren sich heraus, wenn man beachtet, dass der Wille zur Wahrheit, gemäss Nietzsches theoretischen Überlegungen, letzten Endes auf den Glauben an die Gegensätze der Werte zurückzuführen ist. Dies kann nicht nur indirekt über die Bezugnahme zum metaphysischen Glauben erschlossen werden, der in der *fröhlichen Wissenschaft* thematisiert wird. Vielmehr deutet in Jenseits von Gut und Böse Nietzsche selbst eine Relation an, wenn er schreibt: „[...] aus diesem ihrem „Glauben“ [Glaube an die Gegensätzlichkeit der Werte] heraus bemühen sie [die Metaphysiker\*innen] sich um ihr „Wissen“, um Etwas, das feierlich am Ende als „die Wahrheit“ getauft wird“ (KSA 5, 17). Das „Bemühen“ kann hier als Bezeichnung für den Willen zur Wahrheit verstanden werden. Dass der in Anführungszeichen gesetzte Glaube den Glauben an die Gegensätzlichkeit der Werte bezeichnen muss, ergibt sich aus dem Kontext des Aphorismus. Wie eingangs erwähnt, hat dieser Glaube aus Nietzsches Sicht

höchst bezweifelbare Annahmen, die sich als falsch herausstellen könnten. Würden sich diese tatsächlich als falsch herausstellen, dann hätte das die Bedeutung eines Irrtums, auf dem der Willen zur Wahrheit steht. Doch Nietzsche vermutet nicht nur die Möglichkeit eines solchen Irrtums selbst, sondern auch die Möglichkeit eines hinter diesem Irrtum liegenden Willens zur Täuschung, aus welchem heraus dieser Irrtum entstanden ist und diesen aufrechterhält (KSA 5, 16).

Was im ersten Hauptstück von *Jenseits von Gut und Böse* nur als reine Spekulation präsentiert wird, erscheint im zweiten Hauptstück bereits als in eine feste Behauptung. Dort heisst es nämlich im ersten Aphorismus: „Und erst auf diesem granitnen Grunde von Unwissenheit durfte sich bisher die Wissenschaft erheben, der Wille zum Wissen auf dem Grunde eines viel gewaltigeren Willens, des Willens zum Nicht-wissen, zum Ungewissen, zum Unwahren! Nicht als sein Gegensatz, sondern – als seine Verfeinerung!“ (KSA 5, 41). Mit Recht kann eingewendet werden, dass nicht von einem Willen zur Täuschung, sondern von einem Willen zur Ignoranz und zum Irrtum gesprochen wird, was per se nicht dasselbe ist. Dieser Zweifel wird mit Blick auf Aphorismus 230 im siebten Hauptstück ein Stück weit sogar bestätigt, in welchem Nietzsche den „Trieb“ und „Entschluss zur Unwissenheit“ von der „Bereitwilligkeit des Geistes andere Geister zu täuschen“ unterscheidet (KSA 5, 167). Andererseits sollte diese Unterscheidung nicht zu strikt verstanden werden, denn bemerkenswerterweise versteht Nietzsche das Eine wie das Andere als Ausdrucksformen eines Willens, den er als Wille zur Vereinfachung beschreibt (KSA 5, 167-168). Auf die Bedeutung des Willens zur Vereinfachung kann im Rahmen dieser Arbeit nicht eingegangen werden. Doch soll zumindest auf folgende grundlegende Verständnisproblematik hingewiesen werden, die sich im Kontext der Frage nach dem Ursprung des Willens zur Wahrheit ergibt: Wenn der Wille zum Wissen, welcher gleichzeitig der Wille zur Wahrheit ist, mit dem Willen zum Nicht-Wissen als dessen Verfeinerung identifiziert werden kann, und diese Willensformen wiederum mit dem sogenannten Willen zur Einfachheit identifiziert werden können, wie ist es dann zu erklären, dass nach Aphorismus 230 dem Willen zum Einfachen, der „Hang des Erkennenden [...] entgegenwirke“? (KSA 5, 168). Denn es spricht vorerst sehr wenig dagegen und viel mehr dafür, dass der Hang des Erkennenden mit dem Willen zur Wahrheit identifiziert werden kann. Doch insofern dieser Hang dem Willen zur Einfachheit entgegenwirkt, kann hier nicht das Eine als das Andere aufgefasst werden. Von weiterem Interesse im Kontext

einer Auseinandersetzung mit Nietzsches Deutung des Willens zur Wahrheit wäre die Frage, weshalb Nietzsche den Hang des Erkennenden, als „Grausamkeit des intellektuellen Gewissens und Geschmacks“ charakterisiert. Zum Abschluss dieser Untersuchung soll darauf hingewiesen werden, dass Nietzsche selbst hinsichtlich der Frage, warum nach Erkenntnis gestrebt wird, zugesteht, keine befriedigendere Antwort gefunden zu haben, als die, dass das Erkennen eine Aufgabe sei, welche sich der Mensch selbst zur Aufgabe mache (KSA 5, 169-170). Dass der Wille zur Wahrheit Nietzsche zufolge letztendlich im Unbewussten gründet, wird auch zu Beginn in *Jenseits von Gut und Böse* angedeutet: „Was in uns will eigentlich „zur Wahrheit“?“ (KSA 5, 15). Man beachte daran an- und damit auch abschliessend, Nietzsches enorme Beimessung der Bedeutsamkeit, der in dieser Frage geäusserten Ursprungsproblematik, die mit der Wertproblematik einhergeht, gegen Ende der *Genealogie der Moral*: „welchen Sinn hätte UNSER ganzes Sein, wenn nicht den, dass in uns jener Wille zur Wahrheit sich selbst ALS PROBLEM zum Bewusstsein gekommen wäre?... An diesem Sich-bewusst-werden des Willens zur Wahrheit geht von nun an – daran ist kein Zweifel – die Moral zugrunde“ (KSA 5, 410).

## Schlusswort

Die Resultate dieser exegetischen Untersuchung lassen sich wie folgt zusammenfassen: Der Wille zur Wahrheit wird von Nietzsche als unbedingt charakterisiert. Nicht, weil er diesen als Pflicht auffasst, sondern weil dem Willen zur Wahrheit eine Wertschätzung zugrunde liegt, die der Wahrheit gegenüber dem Irrtum einen absoluten, unverhandelbaren Wert beimisst. Diese Wertschätzung, die Nietzsche unter anderem als „Glauben an den metaphysischen Wert der Wahrheit“ bezeichnet, wird zugleich durch den Willen zur Wahrheit zum Ausdruck gebracht. Der Glaube an den metaphysischen Wert der Wahrheit ist auf den metaphysischen Glauben zurückzuführen, dass die Wahrheit in einer zur Lebenswelt entgegengesetzten Welt ihren Sitz hat. Insofern kann die durch den Willen zur Wahrheit geäusserte Wertungsweise mit der Wertung des Lebens durch das asketische Ideal selbst identifiziert werden, da dieser die Vorstellung zu eigen ist, dass die Lebenswelt in einer logisch-ausschliessenden und also gegensätzlichen Beziehung zu einer metaphysischen Welt steht. Insbesondere werden die Werte der beiden Daseinswelten in eine gegensätzliche Beziehung zueinander gestellt, sodass der Wille zur Wahrheit als Bejahung der metaphysischen Welt von Nietzsche zugleich als Verneinung der Lebenswelt verstanden wird. Nietzsche zieht die-

se Vorstellungen in Zweifel, auf denen, seiner Theorie zufolge, der unbedingte Wille zur Wahrheit fußt, und spekuliert ausgehend davon über die Möglichkeit, dass der Wille zur Wahrheit sich erst auf Basis von Irrtum, Täuschung und Unwissenheit erhebt.

## Literatur

### Primärliteratur

- Nietzsche, Friedrich. *Kritische Studienausgabe* 3. G. Colli, M. Montinari (Hrsg.). 2. durchgesehene Auflage. De Gruyter, 1999. [KSA 3]
- . *Kritische Studienausgabe* 5. G. Colli, M. Montinari (Hrsg.). 2. durchgesehene Auflage. De Gruyter, 1999. [KSA 5]
- . *Kritische Studienausgabe* 6. G. Colli, M. Montinari (Hrsg.). 2. durchgesehene Auflage. De Gruyter, 1999. [KSA 6]

### Sekundärliteratur

- Larmore, Charles (2004), „Der Wille zur Wahrheit“, in: Höffe, Otfried (Hrsg.): *Friedrich Nietzsche. Genealogie der Moral*, Berlin: Akademie Verlag, 163-167.
- Kersting, W.: „Pflicht.“ In *Historisches Wörterbuch der Philosophie*, Bd. 7. Darmstadt 2019, Sp. 432.

---

**Khalil Staubli (22)** studiert Philosophie und Geschichte im Bachelor. Er interessiert sich unter anderem für Metaethik, Handlungstheorie und politische Philosophie, sowie Sprach-, Kultur- und Medienphilosophie. Sein Interesse gilt insbesondere aber auch der Schnittstelle und dem Verhältnis zwischen Philosophie und Psychologie.



