Dirk Franken: Against Foundationalism about Persistence-Conditions ................................................................. 1

Holger Leerhoff: Worrall zu Theorienunterbestimmtheit und Strukturenrealismus: Wirklich kein Problem? ...................... 27

Nathan G. Wood: Deploying Racist Soldiers: A critical take on the ‘right intention’ requirement of Just War Theory ............ 53

Walter Veit: Procreative Beneficence and Genetic Enhancement 75

EDITORIAL

KRITERION – Journal of Philosophy is a forum for contributions in any field of analytic philosophy. We welcome submissions of previously unpublished papers, not under consideration for publication anywhere else. Submissions are reviewed in double-blind peer review mode. Contributions should meet the following conditions:

1. The content must be philosophical.
2. The language must be intelligible to a broader readership.
3. The contribution must contain a traceable argumentation.

The length should be between 4000 and 8000 words. Only contributions in English (preferred) and German are accepted.

IMPRESSUM

EDITORS-IN-CHIEF: Christian J. Feldbacher-Escamilla, Alexander Gebharter

EDITORIAL BOARD: Albert J. J. Anglberger, Laurenz Hudetz, Pascale Lötscher, Christian Wallmann

ADDRESS: Franziskanergasse 1, 5020 Salzburg, Austria.

E-MAIL: editor@kriterion-journal-of-philosophy.org

WEB: http://www.kriterion-journal-of-philosophy.org

INDEXING: KRITERION – Journal of Philosophy is indexed and abstracted by The Philosopher’s Index and EBSCOhost Humanities Source. Information about the journal’s ranking is available at SJR. The journal was also approved of satisfying the ERIH (European Reference Index for the Humanities) criteria: ERIH PLUS.

COPYRIGHT: The copyright remains with the authors.

ISSN: 1019-8288
Against Foundationalism about Persistence-Conditions

Dirk Franken

Abstract
In this paper I will argue against a view that I call foundationalism about persistence-conditions. The core of this view is that composite physical objects have their specific persistence-conditions in virtue of these conditions being fulfilled by the object’s physical constituents at various times. I will provide two arguments – the argument from the possibility of instantaneous objects and the argument from the presence of persistence-conditions – which show that this view is untenable. These arguments will also point towards a more adequate understanding of what it means for an object to have certain persistence-conditions. I will expound this understanding and suggest, on its basis, an unorthodox, hylomorphist account of the persistence-conditions of objects.

Keywords: persistence, persistence-conditions, hylomorphism, coincidence of objects

1 Foundationalism about persistence-conditions

To a first approximation, what I call foundationalism about persistence-conditions is the view that any composite physical object has its specific persistence-conditions in virtue of the fact that its physical parts, at the various times of its (the object’s) existence, are related to each other as required by these persistence-conditions. In other words: that an object has its persistence-conditions in virtue of these conditions somehow being fulfilled by its physical constituents.1

The main aim of the present paper is to show that foundationalism about persistence-conditions is untenable. To this end, I will, in section 2 and 3, present two arguments – the argument from the possibility of instantaneous objects and the argument from the presence of persistence-conditions – which show that this account of how an object gets

1
its persistence-conditions leads to bizarre or even unacceptable consequences. These arguments, however, will not only reveal the inadequacy of foundationalism about persistence-conditions. They will also point towards a more adequate understanding of what it means for an object to have certain persistence-conditions. In section 4 I will expound this understanding and suggest, on its basis, an unorthodox, hylomorphist account of how an object gets its persistence-conditions.

In the remainder of this section I shall introduce the idea of foundationalism about persistence-conditions. Before I start doing so, however, let me make explicit two assumptions on which I will rely throughout this paper. Firstly, I assume that all composite physical objects have persistence-conditions and that most of them actually persist. This assumption is, inter alia, meant to rule out so-called exdurantism or stage-theory, the view that composite physical objects are (i) instantaneous (i.e. without temporal extension beyond the instant at which they exist) and (ii) always exist at the same instant (i.e. do not ‘move’ through time) (see [14] or [31] for elaborated statements). Let me point out, however, that this exclusion of exdurantism is primarily for simplicities sake. I am confident that almost everything I will say in this paper could be made compatible with exdurantism by being said in some more complicated manner. Secondly, I assume that persistence-conditions are kind-specific, that is, that, for any kind of object K, there are specific persistence-conditions such that any object has to have these persistence-conditions in order to be of kind K (see [21], [35], [37] for arguments for this assumption). So, for something to be, say, a statue means for it, between other things, to have statue-like persistence-conditions. I leave it open whether this is because objects have their persistence-conditions in virtue of their being of certain kinds or because objects are of certain kinds partly in virtue of their having certain persistence-conditions.

An instructive way to introduce foundationalism about persistence-conditions is by presenting it as a way to overcome a now familiar misconception of persistence-conditions. The main source of this misconception seems to be an unfortunate, but still common terminology. Instead of ‘persistence-conditions’ philosophers often talk of ‘conditions or criteria of diachronic identity’. This triggers the impression that what is at issue are necessary and sufficient conditions of an object at one time being identical with an object at another time. Let us call this the identity-conception of persistence-conditions. It is manifest, for example, in the following characterization of the problem of personal identity through time by Richard Swinburne: ‘What are the logically necessary and sufficient conditions for a person \(P_2\) at a time \(t_2\) to be the same person
as a person $P_1$ at an earlier time $t_1$? [34, p.223]. As various authors point out, this characterization is defective. For one thing, while there are different kinds of persistence-conditions / conditions of diachronic identity for different kinds of objects, there are not different kinds of identity-relations for different kinds of objects – or so I assume. For another thing, the relation of identity is not only ‘utterly simple and unproblematic’ [21, p.92]: It is the relation in which every entity stands to itself and to nothing else. It is also fundamental, which is to say: Nothing informative can be said about the conditions of its obtaining. (See e.g. [21, p.192f], [25, p.86f] or [27, p.466]). So, if there is any substance to the problem of finding the persistence-conditions / conditions of diachronic identity of certain kinds of objects, which clearly seems to be the case, this problem cannot be understood in the way Swinburne’s formulation suggests: as a problem of giving conditions/criteria of the obtaining of identity.

Beyond this, there is a further, less familiar, problem with Swinburne’s formulation. The formulation strongly suggests that Swinburne starts from the presupposition that, at both times, there is an object of the relevant kind (person, in this case) and then asks under what conditions the object at one time is identical to the object at the other time. In other words: He seems to assume that the question of whether certain persistence-conditions/conditions of diachronic identity are met, arises only under the presupposition that there are, at different times, object(s) of the relevant kind. Such an assumption would commit him to assume that the existence of the relevant object(s) at the relevant times is prior to their fulfilling the relevant persistence-conditions / conditions of diachronic identity. But this can hardly be true. Just think of a possible – even if unlikely – situation in which, by cosmic accident, an intrinsic duplicate of you, at this very moment, suddenly alight from a swamp (see [6]). It seems that, at the moment of its coming into existence, this entity would not be a person because, contrary to you, it would lack the right kind of history. More precisely, it would fail to be a person because it would not stand in what Swinburne calls the relation of diachronic identity to a person at former points of time. If this is true, as I shall assume, it cannot generally be the case that the existence of an object at a time is prior to its fulfillment of its persistence-conditions / the conditions of its diachronic identity.

There is far-reaching agreement that the right way to overcome the erroneous identity-conception of persistence-conditions induced by formulations like Swinburne’s is to understand them not as conditions of the identity of objects of specific kinds, but as part of the conditions of
their \textit{existence}.\footnote{Following this idea, Noonan e.g. proposes to replace, for any kind \textit{K}, the talk of diachronic identity-criteria of \textit{K}s by talk of \textquote{the diachronic criterion of \textit{K}-hood} (see \cite[p.91]{25}).} Foundationalism about persistence-conditions is a straightforward and seemingly natural way to work this proposal out. One of its clearest and most elaborated versions is due to David Lewis (see e.g. \cite{19}, \cite{20}, \cite{21}). He used to present it as applied to the prominent example of persons. After having introduced the term \textquote{I-relation} for \textquote{the relation that holds between the several stages of a single continuant person} \cite[p.21]{20}, he states:

\textit{[. . . \ ] something is a continuant person if and only if it is a maximal \textit{[I]-interrelated aggregate of person-stages}. That is: if and only if it is an aggregate of person-stages, each of which is \textit{[I]-related to all the rest} [. . . ], and it is a proper part of no other such aggregate.} \cite[p.22]{20}

As everyone familiar with Lewis’ ontology of physical objects will realize, when he talks of \textquote{stages of objects}, he has in mind temporal parts of four-dimensional objects. His idea can be put as follows: All concrete physical objects are aggregates of temporal stages. So, for there to be persons at least one such aggregate has to fulfill the conditions of being a person. These latter conditions incorporate conditions of two kinds. One says how the stages of an object has to be intrinsically: They have to be person-stages. The other says how such person-stages have to be interrelated to each other: They have to be (maximal) \textit{I}-interrelated. The second kind of conditions are the persistence-conditions of persons. Hence, the persistence-conditions of persons are part and parcel of the conditions of the existence of persons; and analogously for all other kinds of (ordinary) objects.

Now, as e.g. Noonan makes clear, Lewis’ four-dimensional ontology of objects is not mandatory for foundationalism about persistence-conditions. The latter can be maintained under the presupposition of three-dimensionalism as well (see \cite[pp.89]{25}). To account for this possibility, let me introduce an explicitly neutral terminology. Let a \textit{slice} be a portion of matter at a point of time. Such entities are acceptable for both four- and three-dimensionalists. I shall, furthermore, say that a slice \textit{constitutes} an object at a time iff, at the time, the object occupies exactly the same space as the slice.\footnote{In this terminology, both a four- and a three-dimensionalist can – and will – say that physical objects are constituted by different slices at different times. Then, the idea of foundationalism about persistence-conditions amounts to this: For any kind \textit{K}, the conditions of the existence of a \textit{K} incorporate two kinds.
of conditions: (i) conditions on how the slices constituting a K have to be intrinsically: they have to be K-slices; and (ii) conditions on how K-slices have to be interrelated to each other in order to constitute the same K: there is a relation R such that K-slices have to be (maximal) R-interrelated in order to constitute the same K. Conditions (ii) are the persistence-conditions of Ks. Hence, the persistence-conditions of Ks are part and parcel of the conditions of the existence of Ks.

With regards to the issues to be considered in this paper, the crucial aspect of this proposal is the explanation it offers for an object’s having certain persistence-conditions. If the persistence-conditions of an object are part and parcel of the conditions of the object’s existence, an object has its persistence-conditions in the same sense in which it has its existence-conditions. Conditions have to be in force prior to their exemplification. But prior to their exemplification conditions of the existence of objects cannot be ascribed to particular objects. The only objects to which they can be reasonably ascribed are the object’s that come into being in virtue of their fulfillment. So, the only sense in which these objects can be said to have the relevant existence-conditions is that they come into being in virtue of their fulfillment. According to foundationalism about persistence-conditions, the same has to be said about persistence-conditions. That an object has specific persistence-conditions means that it comes into being in virtue of the fulfillment of these persistence-conditions. (See also figure 1 below for an illustration of this idea.) I regard this as the defining feature of foundationalism about persistence-conditions. Accordingly, it will be the leverage point of my arguments against this view.

This being clarified, we should now have a clear idea of how foundationalism about persistence-conditions avoids the pitfalls of the identity-conception of persistence-conditions. Foundationalism about persistence-conditions replaces this misconception by the idea that, instead of being conditions of identity, an object’s persistence-conditions are conditions of the object’s existence. As a consequence, foundationalism about persistence-conditions is free of the problems mentioned above. Recall: The first problem was that, since there are not different kinds of identity, the idea that persistence-conditions are conditions of identity cannot explain why different kinds of objects have different persistence-conditions. No such problem arises for foundationalism about persistence-conditions. Since different kinds of objects have different existence-conditions, the assumption that an object’s persistence-conditions are part of its existence-conditions easily explains why differ-
ent kinds of objects have different persistence-conditions. The second problem was that the conception of persistence-conditions as conditions of identity fails to explain why substantial and informative things can be said about the persistence-conditions of objects. Again, the foundationalist about persistence-conditions has no problems explaining this. According to her, the question of the persistence-conditions of objects of a kind K is the question how K-slices have to be interrelated in order to constitute a K; and this clearly is a substantial issue.

Moreover, at first glance, foundationalism about persistence-conditions also appears promising as a solution of the final problem with the identity-conception of persistence-conditions, the problem of the false presupposition that the existence of object(s) at times is always prior to the question of their fulfilling the relevant persistence-conditions. Since, according to foundationalism about persistence-conditions, the persistence-conditions of objects are part of the existence-conditions of these objects, it cannot be the objects themselves which fulfill these conditions. After all, prior to the fulfillment of the existence-conditions of an object – of which its persistence-conditions are a part, mind you – there just is no object which might fulfill these or any other conditions. The entities on which it is to fulfill an object’s persistence-conditions are, rather, the slices thereby coming to constitute the object. Consequently, according to foundationalism about persistence-conditions, the existence of object(s) at times is not always prior to the question of their fulfilling the relevant persistence-conditions. Quite the contrary, as we saw, foundationalism about persistence-conditions even implies that there can never be an object of a specific kind without its persistence-conditions being fulfilled, that is, being fulfilled by the slices constituting it.

So, foundationalism about persistence-condition is not without its merits. However, the last point already gives us a hint on where it might go wrong. It seems strange to assume that even the coming into existence of an object is a consequence of the fulfillment of its persistence-conditions. Intuitively, persistence-conditions just do not seem to be (parts of) existence-conditions in this sense. Rather, a natural characterization of persistence-conditions seems to be this: They are the conditions under which an already existing object persists (i.e. continues to exist). But if this characterization is along the right lines, an object’s having certain persistence-conditions can hardly be a matter of these conditions being fulfilled. Instead, it would have to be a matter of these condition’s being in force for this object; and the being in force of conditions is, precisely, prior to their fulfillment. The arguments of
sections 2 and 3 are intended to show that this uneasiness with foundationalism about persistence-conditions is perfectly proper and that the alternative conception just indicated is indeed preferable.

For the purpose of presenting these arguments, the following additional bit of terminology will prove helpful: For any object, the temporally first slice constituting this object shall be called its origin and the conditions which make a slice the origin of an object of a kind K shall be called the originating-conditions for Ks. In order not to beg any question, this latter term is to be understood as neutral regarding the question of which kinds of conditions these originating conditions might be. In particular, it allows for the possibility that the originating-conditions for Ks are nothing but the conditions that (i) the relevant slice is, intrinsically, a K-slice and that (ii) it is not related to an earlier slice in the way required by the persistence-conditions for Ks – which is, of course, what a foundationalist about persistence-conditions would say.

2 The Argument from the Possibility of Instantaneous Objects

The first argument is fairly straightforward. Let us call objects that are constituted by just a single slice instantaneous objects. Since persistence requires the existence at more than one point of time, instantaneous objects are objects that do not persist. It is, however, a conceptual truth that, if an object does not persist, its persistence-conditions are not fulfilled. Hence, the single slice constituting an instantaneous object cannot fulfill this object’s persistence-conditions. This raises a problem for foundationalism about persistence-conditions. According to foundationalism about persistence-conditions, objects have persistence-conditions in virtue of these conditions being fulfilled by the slice(s) constituting them. But, as just indicated, the slice constituting an instantaneous object cannot fulfill any persistence-conditions. So, if there were such objects and foundationalism about persistence-conditions were true, these objects would not have any persistence-conditions. This consequence, however, would contradict the two assumptions I made at the beginning of the paper – the assumptions that (i) all objects have persistence-conditions and that (ii) their having the persistence-conditions they have is part of what makes them the kind of objects they are. Hence, if these assumptions are indeed true – what I assume (but see below) –, the foundationalist about persistence-conditions has to deny the possibility
of instantaneous objects.\textsuperscript{10} In other words: According to foundationalism about persistence-conditions, the fact that a single slice meets the originating-conditions for a certain kind of object is not sufficient for there to be an object of the relevant kind with the persistence-conditions of the relevant kind. For the latter to be the case, there has to be at least one further slice being related to this slice as required by these persistence-conditions.

But things do not seem to be this way. It seems absurd to assume that, for any kind of object, there could not be objects of this kind (i.e. objects with the persistence-conditions specific for this kind) which do not exist long enough for their persistence-conditions to be fulfilled. The more reasonable assumption is that, for all (most, some... ) kinds of objects, it is sufficient for there to be an object of the relevant kind (i.e. an object with the persistence-conditions specific for this kind) if a slice meets the originating-conditions for this kind of objects.

To make the point more vivid, consider the following pair of cases:

\textbf{Case 1:} An artist arranges particles of a suitable kind in a way such that a slice at a point of time $t_1$ meets the originating-conditions for statues. The slices at the later points of time $t_2 \ldots t_n$ are related to this slice in the way required by the persistence-conditions of statues. For short: The slice at $t_1$ is the origin of a statue which continues to exist until $t_n$.

\textbf{Case 2:} At and up to $t_1$, everything is as in Case 1. But in the very next moment (i.e. before a slice takes place which fulfills the persistence-conditions of statues) God intervenes and destroys the arrangement of particles.

Since nothing can be a statue without having the persistence-conditions of statues, foundationalists about persistence-conditions are committed to deny that the artist in Case 2 succeeds in creating a statue. Due to god’s intervention, the would-be statue does not exist long enough for the persistence-conditions of statues to be fulfilled.\textsuperscript{11} This conclusion, however, seems plain wrong. The right thing to say seems to be that, in Case 2, there is a statue, but one that exists for only one moment (i.e. an instantaneous statue).\textsuperscript{12} Any view incompatible with this should be avoided.

Well, as it stands, this argument might look suspiciously simple. So, let us see whether there are any reasonable objections a foundationalist
about persistence-conditions might raise against it? I take it that the fact that, given my initial assumptions, foundationalism about persistence-conditions is incompatible with the possibility of instantaneous objects is undeniable. Thus, there remain two options for the foundationalist about persistence-conditions to escape from the argument from the possibility of instantaneous objects: denying the possibility of instantaneous objects and denying my assumptions. I shall consider both options in turn.

As far as I can see, there are three strategies to contest the possibility of instantaneous objects that have some initial plausibility. First: On might argue that the fulfillment of the originating-conditions of any object need some amount of time and can, thus, never be fulfilled by a single slice. To assess this line of argument, some background is needed. To begin with, it is certainly true that there are physical properties of objects such that if an object has such a property at a time, this is due to the fact that the slice constituting it at the relevant time stands in specific relations to slices (whether or not they constitute the same object) at earlier or later times. For any such property, it is true what the objector says about the properties whose exemplification fulfills the originating conditions of objects: Their exemplification needs time. For our purposes, it makes sense to distinguish further between what we might call, following Hawley (see [14, p.53ff]), historical and lingering properties. Historical properties are properties slices have in virtue of their standing in specific relations to former slices and lingering properties are properties slices have (wholly or partly) in virtue of their standing in specific relations to later slices. This conceptual apparatus allows us to restate the objection more precisely. Apparently, there is no incompatibility between the possibility of instantaneous objects of a kind and the assumption that the fulfillment of the originating-conditions for this kind is a matter of the relevant origins having certain historical properties. So, the claim to which the objector is committed is that the originating conditions for all possible objects require the exemplification of lingering properties. This claim, however, is hardly defensible. Consider, again, the example of statues. I see no particularly plausible candidate for a lingering property whose instantiation by a slice might be required for this slice to be the origin of a statue. But even if I were wrong with respect to this and further examples, this would fall far short of proving the impossibility of objects whose originating conditions do not require the exemplification of lingering properties. And I know of no way to make the latter assumption even approximately plausible. So, the first strategy is rather unpromising.
Second: Another strategy appeals to vagueness. One might argue as follows: For all objects, it is vague at which point of time they begin to exist. This means that, for no object, there is a fact of the matter as to which slice is its origin. However, instantaneous objects are, by definition, constituted by their origins alone. So, for an instantaneous object, there being no fact of the matter as to which slice is its origin is the same as there being no fact of the matter as to whether it exists at all – or, alternatively, whether it is an instantaneous object. Hence, there are no clear cases of instantaneous objects. The prospects of this argument depend, for one thing, on what vagueness is. In broad outline, there are three kinds of accounts of vagueness: linguistic accounts according to which vagueness is a lack of precision of our linguistic representation of the world, epistemic accounts according to which vagueness is a lack of knowledge of the precise boundaries in the world and, finally, worldly accounts according to which vagueness is a real world phenomenon. We need not get into the details of these accounts\textsuperscript{15} to realize that the above argument is unsound if the linguistic or epistemic account is true. On both accounts, vagueness is not a feature of the world itself, but of our representation of the world. If one of them is true, vagueness regarding the beginning of an object does not entail that there is no fact of the matter as to which slice is this object’s origin. It would just entail that we fail to know this fact or fail to represent it precisely. Hence, the diagnosis of vagueness in the linguistic or epistemic sense does nothing to support the conclusion that there are no clear cases of instantaneous objects. Things would be different if vagueness were a real world phenomenon. In this case, vagueness regarding the beginnings of objects would indeed entail that, for the relevant objects, there would be no fact of the matter as to which slices were their origins. The commitment to the worldly account is, however, already a heavy burden for the objector. After all, there are strong arguments against the possibility of vague objects, arguments which convinced a majority of philosophers to deny it (see e.g. [8] and [31, chpt.4.9]). For this reason alone, I might not have too much to fear from the assumption of vagueness. But even if it can be shown that vagueness is a worldly phenomenon and that there are objects with indeterminate origins, this does still not entail what the objector need to show: that there can be no objects with determinate origins. And, again, I know of no way to make the assumption of this impossibility even approximately plausible.

Third: Worries regarding the possibility of instantaneous objects might also arise from the commonly held assumption that time is di-
visible at infinitum. As e.g. Hawley convincingly argues, what I call a slice should be as fine-grained as time itself (see [14, p.48]). So, if time is divisible at infinitum, the slices have no temporal extensions whatsoever. This gives raise to various tricky issues: How can slices without temporal extension add up to (the constitution of) an object with temporal extension? How could even God intervene before a second statue-slice can come into existence? I have to concede that I have no answers to these questions. But, fortunately, all problems arising from the infinite divisibility of time are at least as much the problems of the foundationalist about persistence-conditions as they are mine. After all, the foundationalist about persistence-conditions assumes that (i) objects are constituted by aggregates/successions of slices which have both intrinsic and relational properties and that (ii) an object’s persistence-conditions are a matter of the relevant slices being related to each other in specific ways. By making these assumptions she is also committed to assume that – vagueness aside – there is, for every object, one such slice which is the temporally first one constituting this object. So, it is up to her to make sense of the possibility that this slice is not related to later slices in the way required by the persistence-conditions of the relevant object. If she cannot do this, so much the worse for foundationalism about persistence-conditions. But if she can do this, the challenge posed by the argument from instantaneous objects remains in full force. So, the third attempt to save foundationalism about persistence-conditions by denying the possibility of instantaneous objects fails as well.

What is more, even if one of the above attempts to disprove the possibility of instantaneous objects succeeded, this would not automatically rescue foundationalism about persistence-conditions. Note that none of the alleged reasons against the possibility of instantaneous objects that these attempts invoke are essentially related to the reason for which this possibility is ruled out by foundationalism about persistence-conditions. So, it might still be that foundationalism about persistence-conditions fails because it rules out the possibility of instantaneous objects for the wrong reasons. And, indeed, this suspicion comes naturally to mind. Foundationalism about persistence-conditions rules out the possibility of instantaneous objects because it regards the having of persistence-conditions as a matter of their fulfillment. And this seems to be a mistake regardless of whether, for whatever reason, instantaneous objects are in fact impossible. This suspicion shall be reinforced in what follows.

What about the second option, the option of denying my initial assumptions that all objects have persistence-conditions and that their
having these persistence-conditions is part of what makes them the kind of objects they are? Since these assumptions are among the premises from which I start in this paper, I will not consider this option at length here. At any rate, while the assumptions are, of course, not sacrosanct, it should be clear that denying them would be a high price to pay. Besides that, even this would only refer to another problem: Whether or not these assumptions are true, contrary to what foundationalism about persistence-conditions implies, it does not seem to be the case that they are incompatible with the possibility of instantaneous objects.

3 The Argument from the Presence of Persistence-Conditions

According to foundationalism about persistence-conditions, an object has its persistence-conditions in virtue of these persistence-conditions being fulfilled by the slices constituting the object. As a consequence, an object A and an object B which are constituted by the same plurality of slices have the same persistence-conditions. Moreover, it is to be expected that different pluralities of slices, even if they have some elements in common, almost always fulfill different persistence-conditions. So, as a further consequence of foundationalism about persistence-conditions, if an object A and an object B are constituted by different pluralities of slices which have some elements in common, A and B almost always have different persistence-conditions. Let us say that objects A and B coincide permanently iff $A \neq B$ and A and B are constituted by the same slices; and let us say that objects A and B coincide temporarily iff $A \neq B$ and A and B are constituted by different pluralities of slices which have some elements in common. Since objects with different persistence-conditions cannot be identical, it follows from the above that foundationalism about persistence-conditions implies the possibility of temporary coincidence. Moreover, given the highly plausible assumption that, if objects coincided permanently, their distinctness would always be (at least in part) a matter of a difference in their persistence-conditions, it also follows from the above that foundationalism about persistence-conditions rules out the possibility of permanent coincidence. So, proponents of foundationalism about persistence-conditions are committed to both: the acceptance of temporary coincidence and the rejection of permanent coincidence.
While just this view is happily endorsed by many (see e.g. [13, 21, 24, 26, 31]), closer investigation shows that it leads to rather bizarre consequences. Consider the following pair of cases:

Case 1: An artist molds two lumps of clay of similar size and puts them together in such a way that the resulting lump of clay forms a statue. After some time, the statue breaks into two pieces roughly corresponding to the lumps of clay from which it was originally created.

Case 2: Like case 1, except that the statue does not break apart, but is destroyed by being squashed.

Case 1 is of course Gibbard’s famous case of Lumpl and Goliath (see [13]). Since, in this case, the statue (if there were one) and the lump of clay (if there were one) by which it is constituted would come in and get out of existence at the same time, it would be a case of permanent coincidence if such cases were at all possible. According to foundationalism about persistence-conditions, however, cases of permanent coincidence are not possible. So, proponents of this view have to deny in one or another way that in case 1 there are two objects, a statue and a lump of clay. In case 2, on the other hand, the statue clearly ceases to exist before the lump of clay. The lump of clay, but not the statue, survives the squashing. So, according to foundationalism about persistence-conditions, case 2 is a case of two objects, a statue and a lump of clay, coinciding temporarily.

By assumption, however, both cases do not differ at and up to the point of time at which the statue (or the lump of clay... ) comes into existence. That is, in both cases the slice at this point of time meets the same originating-conditions, which are, in both cases, at least the originating-conditions for statues and the originating-conditions for lumps of clay. So, since the foundationalist about persistence-conditions rejects the possibility of permanent coincidence, she has to deny for case 1 what she accepts for case 2: that the slice meeting both kinds of originating-conditions constitutes two ordinary objects. The consequence is remarkable: A person investigating the relevant slice in one of the cases is, as a matter of principle, unable to settle the question of whether there are before her two ordinary objects, a lump of clay and a statue, or just one. For the answer to this question depends on what will go on in the future. If the statue will break apart, as in case 1, and, thus, the statue and the lump of clay, will cease to exist at once, there always would have been only one ordinary object. If, however, the statue will be just squashed, as in case 2, and, thus, the statue will cease to exist before the
lump of clay, there always would have been two ordinary objects, a lump of clay and a statue. But this seems unacceptable. Is it not evident that at every point of time the number of ordinary objects existing at this point of time is determined by what goes on at and up to this point of time?

Actually, things are even worse. The person investigating the relevant slice is not just necessarily unaware of the number of objects before her, but, to a certain extent, also of their kinds. In case 2, the case of the slice’s constituting two objects, one of these objects is a lump of clay and the other a statue. In case 1, however, the case of the slice’s constituting only one object, this object is either a lump of clay or a statue. Assume for the moment that it is a lump of clay. In this case, there is, consequently, no statue in case 1. As we just saw, however, in case 2 there is a statue. So, since a person investigating the relevant slice does not know whether she is confronted with case 1 or with case 2, she is, as a matter of principle, not in the position to say whether there is a statue before her. (Equally, if the single object in case 1 were a statue, instead of a lump of clay. In this case, she would be unable to say whether there is a lump of clay before her.) But this seems even more unacceptable.17

**Question 1:** Might it not be the case that in case 1 there is one object that is both a statue and a lump of clay (see e.g. [2])?

**Answer:** This could be said only if at least one of ‘statue’ and ‘lump of clay’ were not a sortal term, a term for what kind of thing an object is. For if both are sortal terms they provide different persistence-conditions which implies distinctness of the objects falling under them. Now, it might be that both terms also have non-sortal uses or that one might give them such uses. But this does not change the fact that both terms can be used as sortals; and since this is how they are used in the argument, the argument remains unaffected.

**Question 2:** The conclusion that foundationalism about persistence-conditions is incompatible with the possibility of permanent coincidence rests on the assumption that if two or more objects coincide permanently, their distinctness is, at least in part, a matter of their having different persistence-conditions. Could not the foundationalist about persistence-conditions avoid the conclusion by rejecting this assumption?

**Answer:** Hardly! For the assumption to be false, the distinctness of permanently coinciding physical objects would have to be grounded
in something not only different, but altogether independent from their persistence-conditions. The only remotely plausible candidate for this seems to be their having different modal properties. So, rejecting the assumption amounts to claiming that, if two or more physical objects coincide permanently, they differ in their modal properties without differing in their persistence-conditions. This already suffices to disqualify all familiar proposals of cases of permanent coincidence. For in none of these cases the allegedly different modal properties of the relevant physical objects are independent of their persistence-conditions. And I doubt that cases can be found for which the same is not true.

But even if the claim could be defended, it is highly questionable whether it would be of much use for the foundationalist about persistence-conditions. Since the foundationalist about persistence-conditions assumes that an object has its persistence-conditions in virtue of these persistence-conditions being fulfilled by the slices constituting the object at different times, it is to be expected that she makes a parallel assumption about what we might call transworld identity-conditions: An object has its transworld identity-conditions in virtue of these conditions being fulfilled by the slices constituting the object at different possible worlds. However, just as foundationalism about persistence-conditions entails that, for any object that has certain persistence-conditions, there have to be slices at different points of time which constitute the object by fulfilling its persistence-conditions, this latter assumption entails that, for any object that has certain transworld identity-conditions, there have to be slices at different possible worlds which constitute the object by fulfilling its transworld identity-conditions. I am not sure how many foundationalists about persistence-conditions are prepared to accept such a strong modal realism.

We see that unacceptable consequences follow from the combination of the acceptance of temporary coincidence and the rejection of permanent coincidence. And since temporary coincidence is, as Noonan puts it, ‘about as uncontroversial as anything in philosophy could be’ [24, p.1081], we can conclude that the problem lies in the rejection of permanent coincidence. As we saw, however, this problem has its source in the assumption made by the foundationalist about persistence-conditions that an object has its persistence-conditions in virtue of these persistence-conditions being fulfilled by the slices constituting the object. On this
assumption, persistence-conditions of objects are a function of the pluralities of slices constituting the relevant object. This leads to the unwelcome consequences pointed out in the above argument: that there can be no difference in persistence-conditions between objects constituted by the same slices (i.e. that there can be no permanent coincidence) and that there is no way to come to know an object’s persistence-conditions (and, thus, its kind) without investigating all slices constituting the object, that is, without waiting until the object ceases to exist.

4 Forms to the Rescue!

The argument from the possibility of instantaneous objects and the argument from the presence of persistence-conditions show that foundationalism about persistence-conditions has highly implausible or even unacceptable consequences and is, hence, to be avoided. Beyond that, however, the arguments also show what is wrong with foundationalism about persistence-conditions. As we saw, all problematic consequences follow directly from the assumption that an object has its persistence-conditions in virtue of the fulfillment of these persistence-conditions by the slices thereby coming to constitute the object. So, it is this explanation that is to be given up.

That this is no harm, becomes visible once one considers the question what it means for an object to have certain persistence-conditions in an unbiased way. At the end of section 1 I already stated what strikes me as the most natural characterization of persistence-conditions: Persistence-conditions are the conditions which have to be fulfilled for an already existing object to persist (i.e. to continue to exist). As I also pointed out, this characterization is incompatible with the idea that an object’s having certain persistence-conditions is a matter of these condition’s being fulfilled. Instead, it requires that it has to be a matter of these condition’s being in force for the relevant object; and the being in force of conditions is prior to their fulfillment. Let me call this conception of persistence-conditions – not perfectly impartial, I know – the persistence-conception of persistence-conditions. In what follows, I shall flesh it out a little bit further. To begin with, the persistence-conception of persistence-conditions requires the strict separation of an object’s persistence-conditions from its originating-conditions. The latter, but not the former, are conditions of the object’s coming into existence. To be sure, even according to this new understanding, persistence-conditions are conditions of existence in a sense: They are conditions in
virtue of whose fulfillment already existing objects continue to exist. But they are not the conditions – or part thereof – whose fulfillment brings an object into existence in the first place. Making use of the terminology of origins and originating-conditions we can say: An object’s persistence-conditions are the conditions which have to be met by a slice for this slice to constitute the object (also) constituted by the object’s origin. Saying this, in turn, amounts to the following: Persistence-conditions, while certainly abstract in nature, are not detached from the world of concrete particulars. That they are in force for particular concrete objects, means, rather, that they are features of these objects. In other words: The objects themselves are such that they pose certain conditions onto the world – to wit conditions of their own persistence. Any existing object is such that for a slice to constitute it certain conditions have to be fulfilled by this slice. As a consequence, compared to foundationalism about persistence-conditions, the order of explanation is to be changed. Not an object’s having certain persistence-conditions is explained in terms of the fulfillment of these conditions by the slices constituting the object. Rather, an object’s being constituted by specific slices that fulfill certain persistence-conditions is explained in terms of this object’s having these persistence-conditions. Figure 2 shall illustrate this idea in contrast to the idea of foundationalism about persistence-conditions (Figure 1).
In both figures the downward arrows represent the conditions of existence for a kind, $K$, of objects. In Figure 1 the downward arrow points to the aggregate of all slices constituting the relevant object while, in Figure 2, it points to the object’s origin alone. This illustrates that while, according to foundationalism about persistence-conditions, the coming into being of a particular object is a matter of all its slices meeting certain conditions, according to the persistence-conception of persistence-conditions, it is a matter of only one slice – the object’s origin – meeting certain conditions. Contrary to Figure 1, in Figure 2 there is also a separate arrow representing the object’s persistence-conditions. This illustrates that, in contrast to foundationalism about persistence-conditions, on the persistence-conception of persistence-conditions, persistence-conditions are not (part of) existence-conditions. Moreover, the horizontal direction of the arrow together with its originating in the object’s origin, $O^K$, represents the fact that persistence-conditions are in force only with respect to a specific origin of an object, that is, with respect to an already existing object.

That the persistence-conception of persistence-conditions is indeed preferable to foundationalism about persistence-conditions is shown by the fact that it outstrips the latter with respect to its capability to solve all the problems considered in this paper. For one thing, since it also does not entail the mistake of regarding persistence-conditions as conditions of the obtaining of identity, it faces none of the problems of the identity-conception of persistence-conditions. So, in this regard it is on a par with foundationalism about persistence-conditions. Beyond that, however, it also succeeds where foundationalism about persistence-conditions fails: It affords the correct descriptions of the cases considered in the above arguments. According to the persistence-conception of persistence-conditions, the fact that the artifact in case 2 of the argument from the possibility of instantaneous objects is destroyed immediately after its coming into existence does not prevent this artefact from having the persistence-conditions of statues and, thus, from being a statue. For, if having certain persistence-conditions is not a matter of these persistence-conditions being fulfilled, there is no problem with the assumption that even an instantaneous object has certain persistence-conditions. By the same token, the persistence-conception of persistence-conditions does not entail the impossibility of permanent coincidence which gave rise to the unacceptable conclusion revealed by the argument from the presence of persistence-conditions. If an object’s persistence-conditions are not a function of the slices constituting it, nothing rules
it out that two or more objects have distinct persistence-conditions even though they are constituted by the same slices.

However, the persistence-conception of persistence-conditions just tells us what it means for an object to have certain persistence-conditions. What we do not have as yet, is an alternative answer to the question what makes it the case that an object has the persistence-conditions it has. Let me conclude the paper by offering a cautious suggestion on how such an answer might look like. To begin with, since, according to the persistence-conception of persistence-conditions, an object has its persistence-conditions from the first moment of its existence, its having these persistence-conditions can hardly be grounded in its physical constituents. How could the slice that is an object’s origin bring it about that, for this object, certain persistence-conditions are in force? The suspicion that this is impossible is reinforced by the fact that it seems perfectly possible that two kinds of objects differ in their persistence-conditions while they have the same originating-conditions. In such a case, the object’s persistence-conditions could not be grounded in their origins, for, since the origin of an object of one of these kinds would, necessarily, also be the origin of an object of the other kind, there would not be a difference in the relevant object’s origins that might explain the difference in their persistence-conditions.

What might be an alternative answer? Recall that, in my explanation of what it means for an object to have certain persistence-conditions, I made use of the phrase ‘objects pose conditions of their own persistence onto the world.’ My proposal relies on taking this phrase literally. That is, I claim that physical objects really pose conditions of their own persistence onto the world. How do they do this? Well, the capability of posing conditions onto the world is usually taken to be a privilege of abstract entities, like concepts or propositions. Given this, the following proposal suggests itself: objects have this capability in virtue of containing such an abstract entity as a proper part. Objects are, in other words, partly intensional (or even hyperintensional) entities (see [7, p.305/6] or [10, p.69] for similar remarks).

This idea is familiar from an Aristotelian view (or family of views) which currently undergoes a kind of revival: so-called hylomorphism. According to hylomorphism, composite objects have two aspects: matter and form. For physical objects, to which we shall restrict our considerations, the matter is something like the aggregate of all slices constituting the object throughout the time of its existence and the form is a kind of principle – sometimes called ‘principle of unity’ (see [16]) – which states
the conditions a slice has to meet in order to be part of the matter of the object. Many contemporary hylomorphists assume, moreover, that an object’s form is itself a proper part of the object, alongside the object’s matter (see e.g. [10] or [18]; for criticism, see [9] or [16]). With this assumption in place, my proposal can be understood as a version of hylomorphism. According to such an understanding, the abstract component posing the object’s persistence-conditions onto the world is the object’s form, whereas the object’s origin together with the slices meeting these conditions are its matter.

But while my proposal can be regarded as a form of hylomorphism, it differs significantly from standard versions of hylomorphism. Proponents of these versions fail to take into account the crucial distinction between originating- and persistence-conditions. They think of the form of an object as a single bunch of conditions which are to be met (individually and collectively) by all slices constituting the object throughout its existence. To oversimplify somewhat, they regard the form as ‘picking out’ the whole matter of the object at once (see [10], [16], [18, chpt.7]). Such an understanding leaves no room for the distinction between originating- and persistence-conditions, for the distinction between conditions which are to be met for the object to come into existence and conditions which are to be met for an already existing object to continue to exist. According to such an understanding, all conditions posed by the form are of the former kind. As a consequence, an object’s having certain persistence-conditions cannot but be understood as a matter of these conditions being fulfilled. Thus, standard hylomorphism misses the real character of persistence-conditions in the very same way it is missed by foundationalism about persistence-conditions: by failing to do justice to the fact that an object’s persistence-conditions are set into force only by the object’s coming into existence. According to my proposal, in contrast, an object does not have its form in virtue of the fact that this form is fulfilled by the slices constituting this object. That an object has a specific form (i.e. a specific abstract component) is, rather, a primitive fact. Accordingly, an object’s form does not have to wait for being fulfilled in order to enter into an object. It can be part of the object as soon as it comes into existence.

This, to be sure, is hardly more than a sketch of a positive proposal. Much work will have to be done to make it into a mature philosophical account; work that has to wait for another occasion.


**Acknowledgements**

I am indebted to two anonymous referees and, in particular, to Ralf Busse for various helpful comments on previous versions of this paper.

**Notes**

1 By the phrase ‘...in virtue of...’ I mean to express the relation commonly known as *grounding*. Grounding is said to be a genuinely metaphysical, non-causal dependence-relation which obtains between a fact and one or more other facts if the former obtains in virtue of the latter (for details see [4]). Whether there really is such a thing as the grounding-relation is a matter of some dispute (see e.g. [5], [23], [36]). I, for my part, assume that this is so (at least for the present paper). Accordingly, most of my uses of phrases like ‘p in virtue of q’, ‘p is prior to q’, ‘p because of q’ and so on are meant as expressing this or closely related relation. (Thanks to an anonymous referee for making me aware of the need to be explicit in this regard!) I am pretty sure, however, that my crucial points would stand even if there were no such thing as grounding. In this case, the relevant points would have to be restated in terms of relations like explanation, supervenience, reduction and so on, that is, in terms of members of the family of relations which the idea of grounding aims to unify.

2 Note that ‘kind’ is used here in a restricted sense. A kind in this sense is only what is expressed by a so-called substance-sortal, a predicate that answers the question what an object most fundamentally is (see [22], [32], [37]). Which predicates fall into this class and, thus, which kinds of objects there are, is, of course, a contentious issue.

3 Let me stress, however, that there is nothing wrong with this way of talking as long as one is aware of the danger of being misled into the relevant misconception (see also [21, p.192ff]).

4 Pace Geach (see [11], [12]).

5 With that said, how can it be that when Swinburne and other proponents of the identity-conception of persistence-conditions consider questions of the diachronic identity-conditions of various kinds of objects, they give (i) different answers for different kinds of objects and (ii) answers that are both informative and substantial (see, next to [34], e.g. [30], [33])? The only plausible answer seems to be that they misapprehend what they are doing. They take themselves to be looking for conditions of the obtaining of identity, but really do something else.

6 From now on, I will return to the more adequate terminology of persistence-conditions.

7 Parallel to ‘the identity-conception of persistence-conditions’ and ‘the persistence-conception of persistence-conditions’ (see section 4), this conception might well be called ‘the existence-conception of persistence-conditions’. Its relation to foundationalism about persistence-conditions is best described as the relation between a concept or conception of something and a theory that works out this concept or conception. Since I will deal with the theory directly, I will not have much use for the suggested denomination.
This latter term is, moreover, intended as neutral regarding the question whether an object is something over and above the slices constituting it at the various times of its existence. To anticipate: The view I will suggest at the end of the paper is one of the views according to which this is not the case. It contains the assumption that objects have abstract forms as additional components next to the slices constituting them. With this in view, it is advisable to have, from the very beginning, a terminology that allows for this possibility. For the characterization of Lewis’ position, to be sure, this possibility has no relevance. According to Lewis and other foundationalists about persistence-conditions, an object at a time is nothing over and above the slice constituting it at that time.

It is not always easy to say who is a foundationalist about persistence-conditions. The reason is that the question on which this depends is something of a blind spot in the debate. Most authors are confident with (i) opting for three- or four-dimensionalism or (ii) saying something about what the persistence-conditions of certain interesting objects (e.g. persons) are. But the question in virtue of what objects have their persistence-conditions is rarely tackled explicitly. However, next to Lewis, Armstrong (see [1]), Heller (see [15]), Quine (see [28]) or Sattig (see [29]) are rather clear examples of foundationalists about persistence-conditions.

In fact, I know of no foundationalist about persistence-conditions who explicitly draws this conclusion. I think, this just shows that there is little awareness of the problems I raise in this paper.

Why draw on God here? Because slices are so ‘small’ that only god may be quick enough to destroy an object before its persistence-conditions are fulfilled by a second slice. Hereby, I rely on the assumption that the existence of God or a being with comparable skills is at least possible.

An anonymous referee raises the following, rather cunning, objection against my statue-case: For being build the statue has to be perceived. But being perceived is precisely one of the properties whose instantiation needs some time, i.e. which cannot be had by an isolated slice. First of all, this objection, at best, undermines my example case, but not the whole argument. The argument rests on the assumption of the possibility of instantaneous objects, not on the illustrative case of the statue. Beyond that, I am not sure whether the objection even undermines this case. It is certainly right that, in order to build a statue, an artist has to perceive the matter from which the statue is built throughout the process of the building of the statue. But why assume that she has to perceive the completed statue (due to its immediate destruction) would not change the fact that, at the relevant moment, there was a statue.

Note, however, that Hawley’s use of these terms is not precisely the same as mine.

Indeed, just this seems to be the case for many objects. Consider statues, the example I utilize for my argument: It seems pretty plausible that for a slice to be the origin of a statue it has to instantiate, next to certain intrinsic properties (sufficient solidity, a size in a certain range...) certain complex and, presumably, highly disjunctive historical properties (being brought about by an artist with certain intentions and so on).

See [17] for a thorough introduction.

The qualification ‘ordinary’ is meant to circumvent the following problem: It is all but uncontentious that the relevant slice constitutes at most two objects. If e.g. Lewsian universalism were true, it would – like any other slice – constitute
innumerable distinct objects: one for any set of (past, future and possible) slices to which it belongs. What is worse, it is not even clear that in this case it would follow from foundationalism that the number of objects constituted by the relevant slice differs in both cases. But this is necessary for the argument to go through. If the domain of objects under consideration is restricted to ordinary objects like lumps of clay and statues (whatever this might mean exactly), no such problem arises. However things might be with respect to objects in general, foundationalism about persistence-condition does imply that the number of ordinary objects constituted by the relevant slice differs in both cases – or so I argue.

17 This argument can as well be presented as a transcendental argument. Take as starting point the assumption that it is possible for us, at least in principle, to come to know via empirical investigation how many objects there are at a point of time and of what kinds they are. Since we can have no empirical knowledge of the future, it is a condition of this possibility that these facts do not depend on the future. Foundationalism about persistence-conditions, however, implies that they do depend on the future. Hence, foundationalism about persistence-conditions is false.

18 Here, I will not consider the specific character of the relevant condition. Presumably, however, it is best regarded as the condition to stand in a certain relation to the origin of the relevant object; whereas this relation is to be understood as entailing something like the presence of a continuous succession of slices from the origin to the slice in question.

Dirk Franken
Westfälische Wilhelms-Universität Münster
Philosophisches Seminar
Domplatz 23
48143 Münster, Germany
<dirk_franken@gmx.de>
References


Abstract
In the debate of scientific realism, the argument from underdetermination of theories by evidence is put forward by the anti-realist side. According to this argument, for any scientific theory rival theories can be found which are equally well supported by the evidence but incompatible with the original theory. Structural realism is a form of realism that limits the realistic belief in the existence of the entities and structures talked about in scientific theories: Only the structures of reality are relevant in an epistemic sense and responsible for the truth and falsehood of theories. In the light of the antirealistic arguments, structural realism is supposed to have clear advantages over other varieties of scientific realism. In particular, Worrall ([21]) argues that structural realism is immune against the argument from underdetermination. I will sketch Worrall’s line of argument in a more systematic manner and conclude that it is based on some problematic preconditions.

Keywords: realism, structural realism, underdetermination, non-empirical criteria

1 Vorbemerkung
Das Argument der Theorienunterbestimmtheit (TUB) nimmt in der Debatte um den wissenschaftlichen Realismus eine wichtige, wenn auch etwas unbestimmte Rolle ein. Einer der Gründe dafür dürfte sein, dass es – im Gegensatz etwa zum No Miracles-Argument – schwer zu greifen ist. Es lässt sich (aus skeptischer Perspektive) so stark machen, dass kaum eine realistische Position den sich daraus ergebenden Ansprüchen genügen kann; andererseits lässt es sich auch so sehr abschwächen, dass fast jeder Realist sich ohne größere Probleme damit arrangieren kann.

Worrall argumentiert in [21] für die These, dass die von ihm vertretene Ausprägung des Strukturenrealismus durch TUB keiner Gefahr


2 Worralls Strukturenrealismus

existieren. Viele Verfechter dieser Position befassen sich auch intensiv mit
der Philosophie der Physik; für sie hat der ontische Strukturenrealismus
den Reiz, dass sich einige der im Zusammenhang mit Quantenobjekten
auftretenden Phänomene mit seiner Hilfe widerspruchsfrei darstellen las-
sen.\textsuperscript{1}

Worralls Argument für die Immunisierung des Strukturenrealismus
gegen TUB stellt nun, wie sich zeigen wird, einige Anforderungen an
die Ausprägung des Strukturenrealismus, welche die Reichweite seines
Arguments stark einschränken. Meiner Einschätzung nach könnten sich
nur Verfechter einiger dem epistemischen Strukturenrealismus zuzu-
rechnenden Positionen auf Worralls Lösung berufen. Wenn ich im Folgenden
verkürzt von \textit{dem} Strukturenrealismus spreche, beziehe ich mich damit
speziell auf Worralls Variante und solche, die dieser in den für sein Ar-
gument relevanten Aspekten nahestehen.

Ein zentrales, oft aber recht vage oder knapp mathematisch beschrie-
benes Element der strukturenrealistischen Positionen ist das der \textit{Struk-
tur}. Worrall ist in diesem Punkt erfreulich klar: Anknüpfend an frühere
wissenschaftstheoretische Debatten geht er von einer theorie- respektive
theoriesystemabhängigen Beobachtungsgrenze aus und knüpft auch an
die einschlägige Trennung von Beobachtungs- und theoretischer Spra-
che an. Für Worralls Konzeption der Struktur einer Theorie und insbe-
sondere für sein Immunisierungsargument spielt auch die Methode der
Ramseyfizierung eine tragende Rolle. Mit diesem 1929 von Ramsey \textsuperscript{16}
beschriebenen Verfahren lassen sich Theorien mit theoriesprachlichen
Termini so umformen, dass sie in reiner Beobachtungssprache formuliert
sind und alle in ihnen vorkommenden direkten sprachlichen Verweise auf
Unbeobachtbares getilgt werden.\textsuperscript{2} Den Grundgedanken gibt Stegmüller
so wieder: „Den Ausgangspunkt bilde eine in ihrem außerlogischen Teil
endlich axiomatisierte interpretierte Theorie. Man verknüpfe die Axi-
ome konjunktiv mit den Zuordnungsregeln zu einem einzigen Satz. Nach
Beseitigung evtl. vorkommender theoretischer Individuenkonstanten er-
setze man alle theoretischen Prädikatkonstanten durch verschiedene freie
Variable und stelle dem so gebildeten Ausdruck ein Präfix voran, beste-
hend aus Existenzquantoren, durch welche alle neu eingeführten Variab-
len gebunden werden.“ \textsuperscript{17, S. 403}

Nun muss man nicht zwangsläufig akzeptieren, dass eine ramseyfi-
zierte Theorie mit ihrer Ursprungstheorie identisch ist. Während sich
zeigen lässt, dass sie sich hinsichtlich ihrer deduktiven Leistungsfähigkeit
nicht unterscheiden (siehe etwa Stegmüller in \textsuperscript{17, S. 409–411}), kann man
durchnaus die Auffassung vertreten, dass bei der Ramseyfizierung etwas
verlorengeht, ein sich nicht in Beobachtungssätzen wiedergebbarer „Be-
deutungsüberschuss“ (siehe [17, S. 413]; auch Worrall in [20, S. 148] und
[21, S. 168]). Zudem muss bedacht werden, dass es die Methode der
Ramseyfizierung nicht gibt: Selbst wenn eine Theorie in axiomatisierter
Form vorliegt und die Zuordnungsregeln vorhanden sind, können diese
in unterschiedlicher Weise (etwa: Reihenfolge) in einen Ramsey-Satz
überführt werden, was unterschiedliche Ramsey-Sätze zur Folge hätte;
dieses Problem ließe sich allerdings technisch ausräumen. Um des Argumentes
Willen sei hier jedoch vorausgesetzt, dass der Ramsey-Satz einer
Theorie eindeutig ist und zudem eine bedeutungserhaltende Formulie-
 rung der Theorie – in Worrals Worten: „a theory’s full cognitive content
is captured by its Ramsey sentence“ [20, S. 147] – darstellt.

Eine ramseyfizierte Theorie ist in reiner Beobachtungssprache formu-
liert, geht aber in ihren Aussagen über das Beobachtbare hinaus: In ihr
finden sich auch Aussagen über Unbeobachtbares, die Theoriestrukturen,
auf die es Worrall ankommt. Auf dieser Grundlage können beim Be-
trachten zweier empirisch äquivalenter, zu einer gegebenen Menge an Be-
obachtungsdaten BT „passender“ Theorien $T$, $T'$ drei Fälle unterschie-
den werden: (1) $T$, $T'$ können nur oberflächlich verschieden sein; in diesem Fall würden sie hinsichtlich der Beobachtungsdaten übereinstimmen und zudem identische Theoriestrukturen aufweisen ($ST = ST'$). (2) $T$, $T'$
können nur marginal verschieden sein; in diesem Fall würden sie hinsicht-
lich der Beobachtungsdaten übereinstimmen, aber leicht unterschiedliche
Theoriestrukturen aufweisen ($ST \neq ST'$ und ST ähnlich ST'); dieser
Fall könnte etwa bei der Weiterentwicklung wissenschaftlicher Theori-
en auftreten und deren evolutionäre Modifikationen auffangen. (3) $T$, $T'$
können sehr verschieden sein; in diesem Fall würden sie zwar hinsichtlich
der Beobachtungsdaten übereinstimmen, aber deutlich unterschiedliche
Theoriestrukturen aufweisen ($ST \neq ST'$ und ST unähnlich ST').

Der die strukturgleichen Theorien betreffende Fall (1) wird von Worrall ausführlich diskutiert. Ein Strukturenrealist worrallscher Cou-
leur würde leugnen, dass es sich in einem solchen Fall bei $T$ und $T'$
überhaupt um unterschiedliche Theorien handelt: Seiner Lesart nach wären die strukturgleichen Theorien $T$ und $T'$ lediglich unter-
schiedliche Formulierungen einer einzigen Theorie, was sich an ihren
übereinstimmenden Ramsey-Sätzen erkennen ließe. Darauf werde ich im
Abschnitt Strukturelle Immunisierung näher eingehen. Fall (2) ist für
einen Strukturenrealisten ebenfalls von größtem Interesse, da Theorien
mit nur leicht unterschiedlichen Theoriestrukturen auch hinsichtlich der
Frage ihrer annähernden Wahrheit deutliche Gemeinsamkeiten aufweisen
dürften. Im Abschnitt *Eine Frage der Wahrheit* werde ich dies detailliert diskutieren. In Fall (3) hingegen sind die Theorierivalen auch für einen Strukturenrealisten echte Theorierivalen und können nicht beide gleichzeitig als (annahernd) wahr betrachtet werden.

3 *Theorienunterbestimmtheit*

Das in der wissenschaftstheoretischen Realismusdebatte neben der Pessimistischen (Meta-)Induktion wichtigste antirealistische Argument ist das der Theorienunterbestimmtheit (TUB). Grob umrissen gibt der Skeptiker darin zu bedenken, dass zu jeder wissenschaftlichen Theorie $T$ eine unendliche Anzahl empirisch äquivalenter und logisch inkompatibler Alternativen $T'$, $T''$, ... existieren. Wenn der Realist nun an die Wahrheit einer dieser der Annahme nach empirisch adäquaten Theorien glaubt – und damit auch an die Existenz der im Rahmen dieser Theorie postulierten Entitäten und/oder Strukturen –, ist er darin nicht gerecht- fertigt: Ebenso gut könnte schließlich eine der möglicherweise bekannten, möglicherweise (noch) unbekannten Alternativen wahr sein. Der Realist täte entsprechend gut daran, so der Skeptiker, die empirisch äquivalenten Theorien schlicht als empirisch adäquat und gleichwertig zu betrachten und sich eines weitergehenden Urteils zu enthalten.

Die aktuelle Diskussion von TUB ist bestenfalls als „unübersichtlich“ zu bezeichnen, was nicht zuletzt daran liegt, dass ganz unterschiedliche Varianten des Arguments diskutiert werden, ohne dass dies im Einzelfall immer deutlich herausgestellt würde. Eine umfassende Systematisierung mit illustrierenden Beispielen findet sich bei Park in [13], der eine Klassifikation der Unterbestimmtheit entlang vierer Dimensionen vorschlägt: (1) permanent vs. transient: Eine transiente Unterbestimmtheit lässt sich durch künftige empirische Beobachtungen auflösen, die dann mit nur einer der rivalisierenden Theorien vereinbar sind; bei permanenter Unterbestimmtheit ist das nicht möglich, die Theorien sind hier empirisch äquivalent. (2) moderat vs. radikal: Moderate Unterbestimmtheit liegt dann vor, wenn sich die Theorierivalen hinsichtlich ihrer Behauptungen über Unbeobachtbares nur marginal unterscheiden; bei radikaler Unterbestimmtheit sind die Unterschiede gravierender. Diese Dimension kann mit dem Konzept annähernder Wahrheit von Theorien in Verbindung gebracht werden. (3) aktuell vs. (nur) möglich: Eine aktuelle Unterbestimmtheit besteht zwischen zwei tatsächlich vorliegenden Theorien, bei einer (nur) möglichen Unterbestimmtheit wird kein tatsächliches Vorliegen der Theorierivalen gefordert – dass es solche Rivalen geben könnte,
wird als ausreichend betrachtet. (4) normal vs. bizarr: Bei der normalen Unterbestimmtheit haben die Theorien den Charakter „normaler“ wissenschaftlicher Theorien, bei der bizarren Unterbestimmtheit werden die Theorierivalen künstlich mittels Algorithmen aus der Ursprungstheorie entwickelt.

Die von Worrall zur Diskussion gestellte Version von TUB müsste in Parks Sinne in der ersten Dimension als permanent klassifiziert werden, in der dritten als (nur) möglich und in der vierten als normal. Er interessiert sich also für „normale“ wissenschaftliche Theorien mit ebensolchen Theorierivalen, die jedoch nicht unbedingt direkt vorliegen müssen. Entscheidend ist jedoch, dass keine künftigen empirischen Beobachtungen eine Entscheidung zwischen den rivalisierenden Theorien ermöglichen würden – sie sind empirisch äquivalent, unterscheiden sich also nicht hinsichtlich der aus ihnen ableitbaren Beobachtungsaussagen. Wie eingangs erwähnt, lassen sich entlang der von Park beschriebenen Dimensionen ganz unterschiedliche Varianten von TUB konstruieren; für die folgende Diskussion sollen die hier getroffenen Festlegungen jedoch als fixiert betrachtet werden.

In der zweiten Dimension von Park betrachtet Worrall eine lediglich moderate Unterbestimmtheit – also etwa das Vorliegen unterschiedlicher, aber strukturell ähnlicher konkurrierender Theorien – als unschädlich für den wissenschaftlichen Realismus; insofern müssten Theorierivalen in seinen Augen radikal unterbestimmt sein, um für den Strukturenrealismus eine Gefahr darzustellen: Während bei der moderaten Unterbestimmtheit noch behauptet werden könnte, dass die rivalisierenden Konkurrenten gleichzeitig (annähernd) wahr sind, wäre dies bei der radikalen Unterbestimmtheit nicht mehr möglich. Im Abschnitt Eine Frage der Wahrheit werde ich auch auf diesen Problemkomplex, der in Worralls Argumentation eine Schlüsselrolle einnimmt, näher eingehen.

4 Empirische Äquivalenz und Adäquatheit neu verstanden

Beobachtungsdaten und die auf ihnen basierenden Auffassungen empirischer Äquivalenz und empirischer Adäquatheit zeichnen sich durch ihre intersubjektive Zugänglichkeit und Objektivität aus, was ihre traditionell hervorgehobene Stellung in der wissenschaftstheoretischen Diskussion der letzten gut 100 Jahre und auch in der jüngeren Debatte des wissenschaftlichen Realismus begründet.


Beobachtungsdaten sind, in ihrer Gänze zumindest theoretisch, einer objektiven Beurteilung zugänglich. Die Adäquatheitsbewertung von Theorien erfolgt der klassischen Auffassung nach dadurch, dass die Vorhersagen der Theorien mit den Beobachtungsdaten abgeglichen werden: Sie treffen zu oder sie treffen nicht zu. Entsprechend können Theorien als empirisch adäquat akzeptiert oder als empirisch nicht adäquat verworfen werden. Worrall möchte auch hier mit der Tradition brechen und das Konzept empirischer Adäquatheit neu verstanden wissen. Er fordert die Berücksichtigung eines zusätzlichen Kriteriums für das Vorliegen von Adäquatheit: Das Zustandekommen der vorhergesagten Beobachtungsdaten aus den (klassisch verstandenen) empirisch adäquaten Theorien soll analysiert und daraus eine Bewertung der Güte der Theorien abgeleitet werden. "... it by no means follows that theoretical systems based on rival core theories can always be developed that are equally 'empirically successful' or equally empirically supported. There may be—indeed there standardly are—good empirical reasons for preferring one of two data
equivalent theoretical systems to the other.“ [21, S. 171]

Rein praktisch kann man sich das so vorstellen, dass ein Ranking der erfolgreichen, in Worralls Terminologie „datenäquivalenten“ Theorierivalen vorgenommen wird – und mit der Auswahl der optimalen datenäquivalenten Theorie ist die Stufe der empirischen Adäquatheit erreicht. Beobachtungsdaten werden durch in Worralls Sinne empirisch adäquate Theorien also nicht nur empirisch adäquat vorhergesagt, sondern überdies auf eine qualitativ hochwertige Weise. Diese sieht er durch das Kriterium der non-adhocness gewährleistet.

Dieses Kriterium ist, wie Worrall einräumt, nur schwer zu fassen. Im Endeffekt geht es darum, dass die beobachtbaren Phänomene in den empirisch adäquaten Theorien „drop out . . . in a completely natural way“ [21, S. 164], während sie in den lediglich empirisch adäquaten Theorien deren Komplexität – etwa durch eine entsprechende Gestaltung oder Anpassung der Hilfshypothesen – quasi erst verursachen: „Prediction properly understood is simply the opposite of accommodation. A piece of evidence e is accommodated within a theoretical system T based on a core theory C by tailoring specific and/or auxiliary assumptions exactly so as to produce such a system that entails e. A datum e' is predicted by a theoretical system just in case it is deductively entailed by that system but was not accommodated within it.“ [21, S. 163]

In einer ähnlichen Weise hat Worrall bereits die „mature sciences“ charakterisiert: „So my suggestion is that, instead of leaving the notion of maturity as conveniently undefined, a realist should take it that a science counts as mature once it has theories within it which are . . . predictive of general types of phenomena, without these phenomena having been ‘written into’ the theory.“ [18, S. 114] – Die inhaltliche Verwandtschaft beider Konzepte ist kaum von der Hand zu weisen. Die Bewertung der Theorien mittels dieses Kriteriums setzt wieder bei den Theoriestrukturen an: Die hochwertige Weise der Vorhersagen der Beobachtungsdaten ist in den Theoriestrukturen verankert.9


Das Kriterium der non-adhocness ist am oft bemühten historischen Beispiel der Weltbilder von Ptolemäus und Kopernikus recht gut nachzuvollziehen. Stark vereinfacht und auf Worralls Position bezogen könnte man deren Theorien der Planetenbewegung isoliert betrachtet zwar beide als empirisch adäquat$^B$ ansehen, aber nur das von Kopernikus zudem als empirisch adäquat$^B_S$: Beide Theorien gleichen sich in den Beobachtungsdaten und sagen die Planetenpositionen korrekt voraus. Während Ptolemäus aber in seinem geozentrischen Weltbild mit der Epizykelttheorie komplexe Zusatzannahmen machen muss, um die gewissermaßen „widerspenstigen“ Beobachtungsdaten der Planeten in seinem System unterzubringen, kommt das heliozentrische Weltbild des Kopernikus (fast) ohne solche Zusatzannahmen aus. Worralls Kriterium der non-adhocness würde in diesem Fall die kopernikanische Theorie als empirisch adäquat$^B_S$ ausweisen und die Epizykelttheorie des Ptolemäus als nicht empirisch adäquat$^B_S$ verwerfen.

Trotz der intuitiven Plausibilität muss festgehalten werden, dass Worralls Kriterium abseits von solch plakativen Beispielen nicht nur recht schwer zu fassen ist – es ist zudem in einer allgemeinen, nicht kontextabhängigen Form kaum klar zu definieren.

5 Strukturelle Immunisierung

Worralls grundlegender Ansatz zur Immunisierung gegen TUB ergibt sich rein formal aus seiner Konzeption von Äquivalenz und Adäquatheit und bildet die Basis für das weitere Vorgehen. Sein Verständnis empirischer Adäquatheit\textsubscript{BS} und empirischer Äquivalenz\textsubscript{BS} aufgreifend können zwei Versionen der Theorienunterbestimmtheit unterschieden werden: Die „klassische“ Version

\textit{TUB-B:} Zu einer gegebenen Menge an Beobachtungsdaten BT gibt es mehrere empirisch adäquate\textsubscript{B} Theorien \textit{T}, \textit{T}', \ldots, die im Lichte von BT empirisch äquivalent\textsubscript{B} und logisch inkompatibel sind

\ldots sowie die modifizierte Version

\textit{TUB-BS:} Zu einer gegebenen Menge an Beobachtungsdaten BT gibt es mehrere empirisch adäquate\textsubscript{BS} Theorien \textit{T}, \textit{T}', \ldots, die im Lichte von BT empirisch äquivalent\textsubscript{BS} und logisch inkompatibel sind\textsuperscript{10}

Fälle von TUB-B gibt es nun in der Tat; die bisher diskutierten Beispiele fallen etwa darunter. Den Strukturenalisten ficht dies jedoch nicht an, da die Theorien aus diesen Beispielen für ihn gar nicht unterbestimmt sind. Unter Berücksichtigung der Strukturkomponente von Theorien ergibt sich vielmehr die modifizierte Version TUB-BS, die die Theoriestruktur sowohl bei der Äquivalenz als auch bei der Adäquatheit berücksichtigt – also deutlich mehr fordert als eine reine empirische Adäquatheit\textsubscript{B} wie bei TUB-B. Die in TUB-BS postulierte logische Inkompatibilität zwischen den empirisch adäquaten\textsubscript{BS} Theorien \textit{T}, \textit{T}', \ldots besteht damit zwar weiterhin für einen „normalen“ Realisten, nicht aber für einen Strukturenalisten:

\begin{itemize}
  \item \textit{T}, \textit{T}' sind in ihrer ramseyfizierten Form \textit{R}(\textit{T}), \textit{R}(\textit{T}') in reiner Beobachtungssprache formuliert, beinhalten aber Aussagen über zwei Sphären: (1) die Sphäre der Beobachtungsdaten \textit{BT}, \textit{BT}' sowie (2) – in Form impliziter Strukturaussagen – über die Sphäre des Unbeobachtbaren \textit{ST}, \textit{ST}';
\end{itemize}
• genau dann, wenn $T$, $T'$ in ihren Aussagen zu beiden Bereichen ($BT=BT'$ und $ST=ST'$) übereinstimmen, sind sie empirisch äquivalent$_{BS}$;

• wenn $T$, $T'$ aber empirisch äquivalent$_{BS}$ sind, handelt es sich bei $T$, $T'$ für einen Strukturenrealisten gar nicht um zwei Theorien, sondern, da sie in dem Fall auch hinsichtlich ihrer Ramsey-Sätze ($R(T)=R(T')$) übereinstimmen, um eine einzige Theorie in lediglich unterschiedlichen Formulierungen – und entsprechend um einen lediglich scheinbaren Fall von Theorienunterbestimmtheit.

TUB – auf diese Weise verstanden – stellt für den Strukturenrealismus also keine Gefahr mehr dar, denn TUB-BS trifft nach Worrall für einen Strukturenrealisten, da die rivalisierenden Theorien für ihn nicht logisch inkompatibel sind, niemals zu. Die Unterschiede der somit nur scheinbaren Theorierivalen sind ein Oberflächenphänomen und verschwinden bei der Ramseyfizierung: „According to structural realism . . . there can indeed be no empirical reason to prefer one of two ‘rival’ theories that are empirically equivalent in the sense discussed; but this is because there is no significant difference between them—they are not genuinely rivals.“ [21, S. 171] Auf dieser Grundlage kommt Worrall zu dem Schluss, dass der Strukturenrealismus, anders als andere Ausprägungen des wissenschaftlichen Realismus, gegen alle relevanten Fälle von Theorienunterbestimmtheit – solche empirisch äquivalenter $BS$ Theorien nämlich – immun ist.\[11\]


An dieser Stelle ist eine Zwischenbewertung hilfreich. Mit der hier beschriebenen Methode, die ich „strukturelle Immunisierung“ nennen möchte, hat Worrall ausbuchstabiert, was den Strukturenrealismus angesichts der Bedrohung durch TUB auf der rein formalen Ebene ausmacht und ihn von anderen Ausprägungen des Realismus unterscheidet: In meinen Worten die Bildung von Äquivalenzklassen strukturgleicher

Der Schlüssel für die behauptete Immunisierung ist nämlich die von Worrall durch die Hintertür eingeführte Berücksichtigung der Strukturkomponente auf der fundamentalen Ebene der empirischen Adäquatheit: Die bei TUB-B diskutierten empirisch äquivalenten$_B$ und strukturverschiedenen Theorierivalen von $T$ werden bei Worralls TUB-BS auf Grundlage seiner empirischen Adäquatheit$_{BS}$ schon vor dem Erreichen der Adäquatheitsebene „ausgefiltert“ – und es sind genau diese Theorierivalen, die für den Strukturenrealismus auch in TUB-BS problematisch wären. Die Immunisierung steht und fällt entsprechend mit Worralls Konzeption empirisch adäquater$_{BS}$ Theorien.

6 Voraussetzungen empirischer Adäquatheit$_{BS}$


1. Die **Eindeutigkeitsbedingung**: $K$ muss für beliebige empirisch adäquate $B$, aber strukturverschiedene $T, T', \ldots$ eine eindeutige Reihefolge ohne gleich gerankte Theorien festlegen – denn es darf nur eine in seinem Sinne optimale, empirisch adäquate $B_S$ Theoriestruktur geben.


Problem pointiert so zusammenfasst: „Simplicity makes theories more likeable but not more likely to be true.“ [12, S. 77] – Und was Newton-Smith hier für die Einfachheit von Theorien feststellt, lässt sich analog auf andere non-empirische Kriterien – etwa Worralls non-adhocness – übertragen. 12 Diese Problematik ist jedoch sehr komplex und kann an dieser Stelle nicht im Detail erörtert werden; eine ausführliche Diskussion (mit einer von der meinen abweichenden Bewertung) findet sich bei Psillos [14, S. 165–169].

Warum aber muss K überhaupt die Wahrheit einer Theorie garantieren? Auch dieser Punkt ist in Zusammenhang mit Worralls Neukonzeption empirischer Adäquatheit zu sehen. Dem etablierten Verständnis nach gibt es für den Realisten praktisch ein Zweistufensystem: Die erste Stufe ist die der Adäquatheit \( B \) von Theorien (auf der, stark simplifiziert, der Skeptiker verharrt); unter der Voraussetzung, dass wahre Theorien immer auch empirisch adäquat \( B \) sind, ist die wahre Theorie auf jeden Fall im Ergebnis dieser ersten, objektiven Selektionsstufe enthalten. Der Realist selektiert nun in einem zweiten Schritt über ein non-empirisches Kriterium aus den empirisch adäquaten \( B \) Theorien diejenige, die er für (annahernd) wahr hält; das Einräumen eines gewissen Restrisikos für einen „Fehltritt“ ist dabei angesichts der Bedrohung durch die Pessimistische (Meta-)Induktion schon die Regel, wird von den Realisten jedoch nicht als gravierendes Problem gesehen. Worralls Adäquatheit \( BS \) vereint demgegenüber diese beiden Stufen in sich und hat den Anspruch, aus allen Theorien mittels des integrierten Kriteriums K – das er damit in gewisser Weise zu einem empirischen Kriterium aufwertet – direkt eine zumindest annähernd wahre Theorie zu selektieren: Eine Vorauswahl adäquater \( B \) Theorien gibt es nicht mehr. Damit ist an Worralls Selektionsverfahren K aber auch ein ungleich höherer Anspruch zu stellen – durch die Integration von K in die empirische Adäquatheit \( BS \) muss die Wahrheitbedingung erfüllt sein, andernfalls wäre die tatsächlich wahre Theorie nämlich nicht adäquat \( BS \)!

Aus der für die strukturelle Immunisierung nötigen Kombination der Eindeutigkeits- und der Wahrheitsbedingung ergeben sich – abhängig von der Forderung nach Wahrheit \( simpliciter \) oder lediglich annähernder Wahrheit – weitere Schwierigkeiten, worauf ich im folgenden Abschnitt eingehen werde.
7 Eine Frage der Wahrheit

Eine der eingangs beschriebenen Dimensionen bei Park [13, S. 121] ist die der moderaten vs. radikalen Unterbestimmtheit, die in diesem Kontext auf den Anspruch des Realisten an den Wahrheitsgrad einer Theorie übertragen werden kann. Während zwei Theorierivalen bei einer moderaten Unterbestimmtheit nur leicht differierende Aussagen über Unbeobachtbares machen – die Theoriestrukturen sich also nur leicht unterscheiden und die Rivalen mithin beide annähernd wahr sein können – sind die Unterschiede bei einer radikalen Unterbestimmtheit gravierender: Hier sind die strukturellen Unterschiede der Theorierivalen so groß, dass nicht beide annähernd wahr sein können.


Refraktion und prismatischen Dispersion. Die Korpuskeltheorie wurde später durch die Lichtäthertheorie (am prominentesten bei Fresnel) abgelöst: An die Stelle der materiellen Partikel, die sich durch einen leeren Raum bewegen, traten hier Wellenbewegungen in einem mechanischen Medium, dem Äther; auf dieser neuen Grundlage ließen sich weitere empirische Phänomene, etwa Interferenz (wie sie sich in den Newtonschen Ringen zeigte), Diffraction und die von Huygens entdeckte Polarisationserklärungen. Die Lichtäthertheorie wiederum ging später gewissermaßen als ein Sonderfall in Maxwells elektromagnetischer Theorie auf, die Optik und Elektrodynamik in sich vereinte. Hierüber konnten empirische Phänomene wie Wechselwirkungen von Licht und Elektromagnetismus erklärt werden. Maxwell ging dabei weiter von der Existenz eines Äthers als mechanischem Träger der elektromagnetischen Wellen aus; er und seine Nachfolger sahen sich aber nicht in der Lage, dessen Existenz experimentell nachzuweisen. Mit der Relativitätstheorie wurde die Idee des Äthers schließlich aufgegeben und das elektromagnetische Feld selbst als die Entität betrachtet, in der sich die wellenartigen Änderungen ausbreiten. Mit der späteren Photonentheorie wurde das Licht in gewisser Weise wieder als aus Teilchen, Lichtquanten, bestehend betrachtet, welche jedoch ganz neuen Gesetzen unterworfen sind. Auf der empirischen Ebene kamen nun Erklärungen etwa für den photoelektrischen Effekt hinzu. Man erkennt deutlich, dass in dieser Theoriekette über die verschiedenen Theoriebrüche hinweg ein stetiger Zuwachs an empirischer Erklärungskraft und Leistungsfähigkeit der Theorien stattfand. Worrall argumentiert nun ([18, S. 116–120]), dass ein Realist hier angesichts der sich insbesondere in den ontologischen Grundannahmen der Theorien offenbarenden gravierenden Änderungen nicht mehr glaubhaft von einer Form nur evolutionärer Theorieentwicklung sprechen kann: Der Skeptiker könnte mit der PMI schlüssig argumentieren, dass in der Theoriekette wieder und wieder hochwertige, prognostisch starke Theorien retrospektiv als grundsätzlich falsch verworfen wurden, diverse revolutionäre Brüche stattfanden, womit der Realist in Bedrängnis geriete. An dieser Stelle setzt Worrall an und macht in der historischen Reihe radikal unterschiedlicher Theorien eine Invariante aus: Er argumentiert, dass der strukturelle respektive mathematische Teil der Theorien bei den revolutionären Sprüngen trotz der jeweiligen ontologischen Modifikationen weitgehend konstant blieb: „There was continuity or accumulation in the shift, but the continuity is one of form or structure, not of content. . . . Structural realism . . . seems to me to offer the only hopeful way of both underwriting the ,no miracles‘ argument and accepting an accurate ac-
count of the extent of theory change in science. Roughly speaking, it seems right to say that Fresnel completely misidentified the nature of light, but nonetheless it is no miracle that his theory enjoyed the empirical predictive success that it did; it is no miracle because Fresnel’s theory, as science later saw it, attributed to light the right structure. “[18, S. 117]” This example of an important theory-change in science certainly appears, then, to exhibit cumulative growth at the structural level combined with radical replacement of the previous ontological ideas. It speaks, then, in favour of a structural realism.“ [18, S. 120] Um es etwas überspitzt ausdrücken: Der Strukturenrealismus macht mit seinem speziellen Blick auf die Natur wissenschaftlichen Theorien, der Betonung der Theoriestrukturen gegenüber der Ontologie, aus (einigen) wissenschaftlichen Revolutionen Evolutionen und entzieht sie damit der PMI.


Auf die Frage der Wahrheit bezogen heißt das, dass für Worrall und seinen Strukturenrealismus solche Theorien annähernd wahr sind, die später – als Teilstruktur möchte man sagen – in einer größeren, wahren (oder doch wahrheitsähnlicheren) Theorie der Theoriekette aufgehen. Er charakterisiert diese Auffassung so: „The only plausible view, then, is that currently accepted theories are likely to prove ‘merely’ approximately true in the same sense as those earlier and now rejected theories count as approximately true from the vantage point of the current theories. . . . While two mutually inconsistent theories cannot of course both be true, they may both be approximately true—that is, both may emerge as (of course different) limiting cases of some further, superior theory.“ [21, S. 160]

Zudem – doch das nur am Rande – ist das von Worrall beschriebene Kriterium des späteren „Aufgehens“ einer Theorie in einer wahrheitsähnlicheren, umfassenderen Theorie nur post hoc feststellbar; insofern stellt sich die Frage, weshalb Worralls K-Kriterium eine irgendwie geartete Garantie für die gegenwärtige Selektion einer (annähernd) wahren Theorie darstellen sollte. Siehe dazu auch van Fraassen in [5], der einen ähnlichen Einwand in Hinblick auf die Frage vorbringt, ob ein Theorieelement, etwa ein Atom oder der Äther, als nichtstrukturelle Entität oder als Struktur betrachtet werden soll – auch dies lässt sich oft erst im historischen Rückblick feststellen: „Is it not a little embarrassing to start with the thesis that what is preserved through scientific revolutions is the structure attributed to nature, and then to have to identify structure by noticing what has been preserved?“ [5, S. 303]

8 Diskussion

Möchte man den Erfolg von Worralls Vorschlag zur Immunisierung des Strukturenrealismus gegen TUB bewerten, bildet seine strukturelle Immunisierung den besten Ausgangspunkt. Das Verfahren funktioniert, wenn man Worralls Konzeption empirischer Adäquatheit_{BS} akzeptiert. Diese basiert jedoch auf einem nicht-empirischen Kriterium, seiner non-adhocness, auf dem folglich ein großes argumentatives Gewicht lastet. In der Debatte wird Adäquatheit üblicherweise als objektives und verlässliches Kriterium betrachtet, auf das aufbauend der Realist dann ein in seinen Augen sicheres, aber dennoch prinzipiell fehlbares Verfahren anschließt, um unter den adäquaten Theorien die (annähernd) wahre Theorie auszuwählen. Worrall hingegen verlegt sein Selektionskriterium
in die Adäquatheit hinein. Dies kann man so verstehen, dass diesem Kriterium damit eine besondere Last auferlegt wird, da der Anspruch der Objektivität und Verlässlichkeit, der mit der Adäquatheit verbunden ist, mit diesem Schritt auf seine non-adhocness übergeht.

Dass Worralls Selektionskriterium die Eindeutigkeits- und die Wahrheitsbedingung erfüllt, ist alles andere als belegt. Ohne erfüllte Eindeutigkeitsbedingung jedoch versagt seine strukturelle Immunisierung; und ohne erfüllte Wahrheitsbedingung möchte man kaum von Adäquatheit sprechen. Zudem geht Worrall mit dem Anspruch auf eine lediglich annähernde Wahrheit der Theorien und seinem Rückgriff auf ein non-empirisches Kriterium zur Theorieauswahl in zwei wichtigen Bereichen Kompromisse ein, die generell, obschon bei realistischen Positionen weit verbreitet, skeptischen Einwänden Tür und Tor öffnen.

Auf der technischen Seite bleibt offen, wie sich Worralls Konzept annähernder Wahrheit mit der strukturellen Immunisierung in Einklang bringen lassen kann – wendet man seine Wahrheitskonzeption an, versagt nämlich die strukturelle Immunisierung, sodass der Strukturenrealismus TUB wieder in vollem Maße ausgesetzt ist. Damit stünde der Strukturenrealismus im Lichte von TUB aber nicht mehr, wie von Worrall behauptet, besser da als alternative realistische Positionen. Das mit seiner speziellen Konzeption annähernder Wahrheit einhergehende post hoc-Problem – dass sich eine mit seinem Verfahren K als empirisch adäquat\textsubscript{BS} selektierte gegenwärtige Theorie erst retrospektiv als tatsächlich empirisch adäquat\textsubscript{BS} und damit (annähernd) wahr erweisen kann – ist demgegenüber eher als sekundäre Kritik anzusehen.


9 Fazit

Es hat sich gezeigt, dass Worralls Argumentation zwar schlüssig und zumindest intuitiv plausibel ist, seine Konzeption empirischer Adäquatheit\textsubscript{BS} – mit der sein Projekt steht und fällt – aber question begging ist und skeptischen Einwänden Tür und Tor öffnet. Eine mit der empirischen Adäquatheit\textsubscript{BS} zwangsläufig einhergehende Beurteilung der Qualität der Theoriestrukturen, die zu einer gangbaren Anwendung seiner strukturellen Immunisierung führen würde, könnte zudem nur auf Grundlage eines non-empirischen Kriteriums erfolgen, das exakt eine Theoriestruktur als empirisch adäquat\textsubscript{BS} ausweist und zudem Garant für die (annähernde) Wahrheit der so selektierten Theorie ist. Worrall hat aber weder gezeigt, dass sein Kriterium der non-adhocness diesen hohen Anforderungen genügt, noch scheint es einen plausiblen Grund für die Annahme zu geben, dass irgendein non-empirisches Kriterium solchen Anforderungen genügen könnte. Das von Worrall propagierte Konzept annähernder Wahrheit führt zudem grundsätzlich zu einem Versagen seiner strukturellen Immunisierung. Damit kann Worralls Ansatz der Immunisierung seiner und verwandter Ausprägungen des Strukturenrealismus gegen das Argument der Theorienunterbestimmtheit jedoch nur als gescheitert betrachtet werden.
Notes

1 Eine sehr gute Übersicht über die aktuell vertretenen Positionen und die historischen Wurzeln findet sich bei Ladyman in [8].

2 Eine ausführliche Diskussion mit historischem Einschlag findet sich bei Psillos in [14, S. 46–67].


7 Motiviert wird dieser Schritt nicht zuletzt durch Worralls Analyse des empirischen Erfolgs von Theorien: Er unterscheidet auch in diesem Kontext zwischen einer empirischen Stützung, die sich lediglich auf Beobachtungsdaten bezieht, und der von ihm propagierten Form empirischer Stützung, bei der neben den Beobachtungsdaten auch die in der ramseyfizierten Form der Theorien impliziten Strukturaussagen Berücksichtigung finden. (Siehe etwa [19, Abschnitt 4].)

8 Schon Ladan und Leplin deuten in [9] einen vergleichbaren Weg an, wenn sie non-empirische Kriterien für die Begutachtung des Erfolgs von Theorien ins Spiel bringen, insbesondere Kohärenzaspeskte [9, S. 461–465]. Auch sie bleiben dabei äußerst vage, was aber im Rahmen der von ihnen vertreten naturalistischen Traditionslinie weniger problematisch ist als bei Worrall.
Eine ausführlichere Darstellung von Worralls non-adhocness findet sich in [19, Abschnitt 4]; dort untersucht er die mit dieser Unterscheidung einhergehenden Möglichkeiten im Kontext der Bestätigungstheorie.


11 Auch Boyd sieht den Wahrheitsanspruch kritisch. Er schreibt dazu: „... theoretical claims are incapable of confirmation or disconfirmation. We may choose the 'simplest model' for 'pragmatic' reasons, but if evidence in science is experimental evidence, then pragmatic standards for theory choice have nothing to do with truth or knowledge. Scientific realism promises theoretical knowledge of the world, where, at best, it can deliver only formal elegance, or computational convenience." [1, S. 44]

12 Weitere Argumente für die Abschwächung des Wahrheitsanspruchs finden sich beispielsweise bei Chakravartty [3, Abschnitt 3.4] und Psillos [14, S. 266].


14 Die in Worralls Sinne annähernd wahren Theorien würden sich demgegenüber an ganz unterschiedlichen Stellen des Theoriefeldes befinden, auch wenn es dabei regionale Häufungen geben dürfte.

Holger Leerhoff
Carl von Ossietzky Universität Oldenburg
Institut für Philosophie
Ammerländer Heerstr. 114–118
26129 Oldenburg, Germany
<holger.leerhoff@uni-oldenburg.de>
Literatur


Deploying Racist Soldiers: A critical take on the ‘right intention’ requirement of Just War Theory

Nathan G. Wood

Abstract
In a recent article Duncan Purves, Ryan Jenkins, and B.J. Strawser\(^1\) argue that in order for a decision in war to be just, or indeed the decision to resort to war to be just, it must be the case that the decision is made for the right reasons. Furthermore, they argue that this requirement holds regardless of how much good is produced by said action. In this essay I argue that their argument is flawed, in that it mistakes what makes an agent morally good for what makes an act morally good. I argue that the main thrust of Purves et al.’s argument in fact undermines the conclusion they wish to draw, and that the reasons for one’s action do not make an in principle difference to the morality of actions in war. I further argue that this position undermines the traditional ad bellum just war constraint of right intention, and that the morality of actions in war is, at core, only concerned with outcomes. I conclude by clarifying that one’s reasons for action do in fact matter when deciding to enter war or kill in war, but only because one’s reasons significantly impact the way in which one acts. The purpose of this paper is to clear the theoretical space by showing why intentions/reasons do not in principle matter when assessing the morality of war (or killing), but this should not be taken as an argument that we should ignore intentions/reasons altogether.

Keywords: Just War Theory, Intentions, Actions

1 Introduction

According to St. Thomas Aquinas, three things are necessary for a war to be just: first, the authority of the prince; second, a just cause; and third, “it is required that those who wage war should have a righteous intent”\(^2\). In contemporary just war theory, the first requirement, right authority, has seen much criticism, as non-state actors and civil uprisings are increasingly seen as legitimate (or at least capable of acting
legitimately in a military capacity).\textsuperscript{3} The second and third requirements have, however, retained their status as cornerstones of what it means for a war to be just.

In this essay I challenge the third requirement, right intention, and argue that within the morality of war, intentions only matter insofar as they affect the actions one chooses to undertake, thereby altering outcomes. I proceed by examining the argument presented by Purves et al. \textsuperscript{14}, which attempts to demonstrate that within the morality of war, an agent’s reasons for action can decisively alter the moral assessment of the actions carried out by said agents. I argue, \textit{pace} Purves et al., that if two agents will carry out exactly the same actions bringing about exactly the same consequences, then those agents’ \textit{actions} must be found equally moral or immoral. This, however, should not be taken as an indictment of intentions altogether. Indeed, I argue that intentions do matter, as they provide the basis for determining the \textit{moral character} of an agent. Put differently, intentions matter morally, but not with respect to the permissibility of actions, only with respect to the moral assessment of an individual’s character. I close by clarifying that, practically speaking, intentions will have an impact on the moral assessment of most acts of killing or war, but only because the intentions driving action generally impact on the actions undertaken; i.e. the intentions themselves do not matter for the moral assessment of acts, but the effect they have on choices and outcomes does. It is also worth highlighting at the outset that the arguments provided only deal with moral assessments in war. As such, the critique of intentions should not be confused with more general debates concerning intentions and morality.

\textbf{2 Just War Theory and Right Intention}

As mentioned above, the requirement that war be undertaken with a right intention is an old one, found in the works of Aquinas and Augustine before him,\textsuperscript{4} and it is still accepted by prominent theorists today.\textsuperscript{5} Right Intention is an \textit{ad bellum} constraint, which maintains that for a war to be just, it must, among other things, be fought with the right intention.

Generally, this is understood as demanding that a war be fought \textit{for the sake of its just cause}.\textsuperscript{6} That is, if my country is invaded by an aggressive neighboring state which is bent on exterminating some of our population, then this provides a just cause for defensive war. Our war would be fought with a right intention if we entered it \textit{for}
the sake of this just cause, and not for some other ulterior motives, like economic advantage or a plan to annex some of our aggressive neighbor’s territories. It is important to point out here that this does not necessarily mean that these other reasons for war must be absent. After all, one can enter a war in order to resist aggression or assist others in doing so, as the United States did in entering World War II, while also welcoming the fact that war will help improve one’s economic situation. Instead, what right intention requires is that the effective motive be securing the just cause, and that any other possible motives be given a lexically lower rank, such that they cannot impinge on the right intention.

However, right intention, while often only listed as an ad bellum requirement for war to be just also comes in an in bello formulation as well. After all, if a war’s being just requires that it be entered for the sake of the just cause, then an action in war’s being just should also be taken for the sake of furthering the just cause. Moreover, in bello right intention is also a cornerstone of the traditional notion of a just war, for as Augustine says, “[t]he desire to do harm, the cruelty of vengeance, an unpeaceable and implacable spirit, the fever of rebellion, the lust to dominate, and similar things: these are rightly condemned in war”. What in bello right intention demands is that each soldier and commander make his or her decisions with the intent of securing or furthering some moral goal. However, it does not necessarily mean that each soldier and commander actually in fact furthers a moral goal, because the war they are fighting may lack justification altogether. Right intention is a subjective requirement, concerned only with what is inside an agent (in that it concerns mental states and attitudes), and so it may be satisfied even by agents who fight patently unjust fights. All that is required is that they have a personal intent to do good or promote good, even if they are mistaken.

For example, suppose one of Rommel’s soldiers firmly believes that Germany was wronged in the Treaty of Versailles, and that Hitler’s war is thereby justified as a means to redress that past wrong. We may roundly reject this claim, but the soldier believes it, and has some reasons to substantiate his position (even if they are weak or false reasons). Based on this “just cause”, the soldier joins the Wehrmacht and goes out to fight, with the intent of helping to redress that past wrong. In this case, the soldier acts for the sake of furthering a just cause, even though there is in fact no just cause to be furthered.

This may seem counterintuitive at first, but it becomes clear if we concentrate on the fact that the requirement is about intentions, and
not *justifications*. Rommel’s soldier fights with a good intent, but his ignorance or misapprehension of the facts makes him unjustified nonetheless.\(^\text{10}\) In this way, right intention can come apart from just cause or justified action, because the two deal with different categories of things: the former concerns internal mental states, whereas the latter deals with facts and outcomes as well.

So, according to the requirement of right intention, for a war to be *ad bellum* just, it must be the case that the war is entered *for the sake of securing the just cause*. For an action in war to be just, it must be the case that the agent acts *for the sake of furthering a just cause*. Put differently, a just war is one entered *with the intent of securing justice*; a just act in war is one which is carried out *with the intent of furthering justice*. However, neither of these requirements imply that the war or act really are just. All that matters for these requirements are the intentions and expectations internal to the agent. Whether those intentions and expectations track to anything true or morally good in the world is another matter, and one which is captured by other just war constraints (e.g. just cause, proportionality, necessity).

A final point worth noting is that the *ad bellum* and *in bello* formulations of right intention are structurally identical. Both say that an agent (either a state or an individual soldier) is required to have the right intentions (either securing or promoting what the agent determines to be just) when doing what is normally a forbidden act (entering war or killing others). If this is satisfied, along with all other just war criteria, then the act may be deemed permissible. Given that the two conditions are identical in form and content, and differ only in scale, any problems found with one formulation will imply problems for the other as well. This point will become important in the arguments to follow.

### 3 Acting for the Right Reasons

Now, with right intention and its *in bello* variant spelled out, we can begin looking more closely at the argument provided by Purves et al. Their argument focuses on the morality of using autonomous weapons in war, but in making their case they present a more general claim about how reasons affect the morality of actions in war. The core of their argument is the thought experiment *Racist Soldier*, which is set up as follows:

Imagine a racist man who viscerally hates all people of a certain ethnicity and longs to murder them, but he knows he
would not be able to get away with this under normal conditions. It then comes about that the nation-state of which this man is a citizen has a just cause for war: they are defending themselves from invasion by an aggressive, neighboring state. It so happens that this invading state’s population is primarily composed of the ethnicity that the racist man hates. The racist man joins the army and eagerly goes to war, where he proceeds to kill scores of enemy soldiers of the ethnicity he so hates. Assume that he abides by the jus in bello rules of combatant distinction and proportionality, yet not for moral reasons. Rather, the reason for every enemy soldier he kills is his vile, racist intent.  

Purves et al. contend that given the state of this man’s intentions, we have a strong moral reason to not deploy him to the front, because of the fact that he will be killing for such heinous reasons. More than this, they contend that “if we had a choice between deploying either Racist Soldier or another soldier who would not kill for such reasons, and both would accomplish the military objective, we would have a strong moral reason to choose the nonracist soldier.”

Now, it is important to make clear exactly what Purves et al. are claiming, as the thought-experiment conjures up a number of emotional responses and intuitions. First of all, the point is not that the racist soldier kills more people, or innocent people, or uses means which are unnecessarily harmful. Obviously, any of these things would present a problem for deploying such a man, but these various actions would be problematic on a number of just war theoretic grounds. Furthermore, those actions would be condemned on both strictly consequentialist and strictly deontological grounds, and so the example would not help illuminate the supposed moral value of acting for the right reasons. What the example maintains is that the racist soldier acts in precisely the same way as a non-racist soldier would act. The things he does and the outcomes he brings about are identical with those that any other soldier in his position would bring about, even if he were replaced by the most virtuous soldier available. The problem then is not what he does or even how he does it, but rather why he does it.

Second, Purves et al. are arguing not just that the racist soldier acts in a way which is morally problematic, but also that those who send him into combat act in a way which is morally problematic. After all, the claim is that if we were choosing between racist soldier or some other soldier, “we would have a strong moral reason to choose the nonracist
soldier”. The fact that “we... have a strong moral reason” not to send him means that his racist intent changes the moral decision that we make as well.

Finally, it is crucial to note that the point Purves et al. are arguing for is not that the character of the racist soldier is morally flawed (though I am sure they would agree to this), but rather that his actions are. It is obvious that there are moral deficiencies to the character of the racist soldier, since he is, ex hypothesi, racist. However, Purves et al. want to argue for more than this. They maintain that there is something wrong with us sending him to fight, and also with his fighting, because “he is acting for the wrong reasons”. If their argument was only meant to indicate some flaw in his character, then they could provide the title of the thought experiment – Racist Soldier – and leave it at that. However, they aim to show more; that his racism affects the moral assessment of his actions as well. More than this, they argue that his actions are wrong, even though they are the same actions another nonracist soldier would be carrying out. Given this, it is clear that the argument of Purves et al. is meant to show that two identical actions with identical outcomes can have different moral statuses, simply because the agents had different reasons for carrying out those actions.

4 Right Actions - Right Agents

However, have Purves et al. gotten things right? Is it really true that the intentions an agent has when acting affect the moral status of the action? More importantly, can two actions with identical outcomes really be judged to have different moral statuses? I argue that the answer to all of these questions is “no”.

Let us proceed by examining where intentions matter most, and then seeing whether this can help us to understand the role, or lack thereof, of intentions in evaluating actions. It should be uncontroversial that intentions clearly and significantly impact our moral assessments of an agent’s character. If there are two people who do the same thing, but do so for radically different reasons, there may be disagreement about whether or not both actions have the same moral value or status, but all would agree that the moral character of these two agents differ greatly. To see this, let us briefly consider an example.

Suppose that two men see a young boy floundering in the waves and both rush in to save the him. The first man does so simply because he saw a child drowning and knows he has a moral duty to save drowning
children (even if it means the copy of “Famine, Affluence, and Morality” stuffed into his pants pocket is ruined). The second man, however, only swims out in the hopes that the boy’s grateful parents will give him some sort of reward. He is what we might call a “bounty lifeguard”.\(^{16}\) Now, I take it that most would say both men act correctly, i.e. they perform the right action, though I am sure there are some who might quibble about that judgment. However, one thing that is obvious in this case is that the second man is not nearly as good as the first. Put in less colloquial terms, he demonstrates a flawed moral character, because his only reason for action is personal gain, and moral considerations hold no motivating power for him, even when he sees a child drowning and could save that child with virtually no negative consequences for himself (other than having wet clothes for a bit).

So, intentions definitely do matter morally, at least with respect to the evaluation of an individual’s moral character. Acting for the right reasons demonstrates a well developed moral sense, and is to be praised, while acting for the wrong reasons demonstrates an atrophied or even absent moral sense, and is to be blamed or even perhaps punished.

What of the actions themselves though? How are we to judge these given differences in the reasons which motivate them? Let us again consider a pair of cases, what I will call the noble kamikaze and the sadistic marine. The noble kamikaze fights in WWII for the imperial Japanese army out of a sense of honor, loyalty, duty to country, and so forth. Put shortly, he fights for whatever reasons the reader finds to be good civic/moral traits. His fighting is helping to sustain a war machine that has ravaged China, Southeast Asia, and the Pacific Islands, but he fights for good reasons. The sadistic marine, on the other hand, fights because he gets a kick out of hurting others. Since he was a small child, he has always loved hurting others, and the opportunity to crawl through the jungles with a rifle and little oversight gives him exactly what he has always wanted. However, in fighting, he is helping to thwart tyranny, liberate conquered peoples, and halt grave human rights abuses in Southeast Asia and China.

In these two cases, there are three things we can already say with confidence and without (great) controversy. First, the noble kamikaze acts wrongly, because, no matter what his intentions are, he is furthering grave injustice by assisting in the wholesale subjugation and destruction of entire populations.\(^{17}\) In fact, we need know nothing about his intentions to make this judgment, because his actions are already in violation of a number of other criteria of just war theory.\(^{18}\) Second, despite the
wrongness of his actions, the *noble kamikaze* does have a good intention, and on this ground should be judged as having a *good moral character*, or at the very least, as having a better moral character than agents who fight for self-interested or possibly even sadistic reasons. He is wrong in thinking that it is acceptable to fight for Japan, but he just wants to do what is right. The indoctrination he received as a child may have hindered his ability to recognize what *is* objectively right, but one cannot doubt that that is his goal. This renders his character *morally good*, or at least better than the character of agents without such motivations and intentions.\(^{19}\) The third thing we can say with certainty is that the *sadistic marine* displays a *bad moral character*. He acts for no other reason than to harm others, simply because he finds this amusing. Such sadism shows one of the very worst characters we can imagine, and is something to be roundly condemned.

There is still one thing we have not spoken to though: the actions of the *sadistic marine*. We know he is bad, but are his actions bad as well? In fighting he contributes to the liberation of many people, helps to bring down a regime which is tyrannical not just to those it conquers, but to its own people as well, and he assists in the halting of genocide and grave rights abuses throughout Southeast Asia and China. These are all undoubtedly good things, and good things of a significant moral value.*^{20}\) However, he only does these things because it allows him to harm others without fear of punishment. Does this make the actions *themselves* morally bad?

I believe not. It would certainly be better if he had a good intention while doing these things, but the mere presence of a bad one cannot negate all of the good these actions produce. To see this more fully, consider the following example:

*Ignorant Racist Politician*: Suppose there is a politician who wants to keep minorities down, and ignorantly believes the best way to do this is to provide them with many social goods. He thinks that having all these things given to them will make them dependent on the government, and thereby lower their self-reliance and long-term welfare. However, as would be expected, the presence of better schools, stronger infrastructure, and better medical care all conspire to improve the welfare and opportunities of these minorities, thwarting the politician’s plan.

Now, this scenario is structurally similar to *racist soldier* (but not identical). Both act in ways which produce goods, but only act based on
bad reasons. The main difference is that racist soldier imposes harms based on bad reasons, whereas ignorant racist politician distributes goods based on bad reasons. However, both effect a net positive state of affairs via their actions, and both base their actions on racist sentiments. I do not believe anyone would honestly or sincerely say that the ignorant racist politician has done something bad. That is to say, his actions are good ones. His character is not good, given that his main aim is to harm minorities, but his ignorance leads him to consistently help those he so hates. In such a case, his motives will only matter to us insofar as we are concerned with his moral character or believe that those motives will impact his future decisions. Therefore, we may seek to replace ignorant racist politician with another politician who is not racist, but we will not do this because the former’s actions were bad or wrong. In fact, all of his actions were good ones which improved the lot of some of society’s worst off. No, we will seek to replace him because we expect him to do bad in the future, because of his bad intents or bad reasons for action. This, however, says nothing regarding the actions he has already carried out. Ignorant racist politician is simply a bad man who has done good things accidentally.

Before moving on, it may be worthwhile to consider one final argument against the position of Purves et al., and one which does not rely quite as heavily on intuitions and constructed thought-experiments. Let us return to racist soldier, but restate the main elements in a slightly more schematic fashion.

Racist Soldier (Revised): Suppose we have two soldiers, one who is racist and the other who is not. Call them $S_R$ and $S_{NR}$, respectively. If we send $S_R$ to war he will do $X$. If we send $S_{NR}$ to war he will do $X$. The actions they will perform ($X$) are identical and the results will likewise be identical.

By hypothesis, Purves et al. claim that the $X$ performed by $S_{NR}$ would be morally right. This however, is by their own admission the exact same thing that would be performed by $S_R$. Therefore, it must also be morally right for him to do it, even if he is doing it for some other reason. There is nothing that distinguishes one action $X$ from another, so there cannot be any moral difference between them. Put differently, if you and I both do the same thing, and what I do is right, then what you do must be right as well, because of the fact that we both do the same thing.

Furthermore, we can say this and still maintain that there is some truth to the sentiment that there is something less good or perhaps
morally problematic about sending $S_R$ to war instead of $S_{NR}$. However, the truth behind this sentiment has nothing to do with actions. Rather, it is grounded in the fact that, all things considered, we prefer good moral characters over bad ones. This, however, does not entail that the actions performed because of a bad character are inherently bad. As such, intentions do not, in principle, matter for the evaluation of acts.

So, an act in war can be morally right even if performed for the worst of reasons. This is because reasons do not affect the morality of acts, rather they affect the morality of agents. Two identical acts must have identical moral statuses. The agents who perform those acts, however, may have very different moral statuses because of the reasons they had for acting. In this way, reasons will matter morally, just not with respect to acts.²¹

5 Re-examining Right Intention

So far we have been examining individual acts performed by individual agents who possess some particular moral characters. We have argued that the intentions of an individual will not affect the moral assessment of any acts that individual carries out, in virtue of the fact that intentions only matter for the assessment of moral character. What, if anything, can this tell us about the traditional just war requirement of right intention?

Recall that right intention is the demand that war be fought for the sake of its just cause. It is an ad bellum requirement which must be fulfilled in order for a state’s decision to go to war to be just.²² This, however, implies that right intention is a requirement on actions, because the decision to go to war is an action which a state carries out. More than this, failure to abide by right intention does not imply that a state is wrong, but rather that the state acts wrongly by waging a war without having the appropriate motives. This clearly shows that the requirement is concerned with actions.

However, if what has been argued above holds any truth, then right intention cannot be maintained, because the intentions of a state do not matter for the moral assessments of that state’s actions. Just as an individual agent may act for bad reasons but do a good thing, a state, which is nothing more than a collective agent, may enter a war for bad reasons but do a good thing. For example, a state may enter war purely for territorial gain, but in doing so halt an ongoing genocide in the region it is annexing. In such a case, we clearly have a reason
to criticize that state, morally speaking, given that it values territorial expansion over the preservation of human lives (this is evident in that its motive is expansion rather than other-defense), but this criticism is aimed at the state itself, not its war. A war which does good is a good thing to do, and if there is also a just cause for war, then it is the right thing to do. This is the case irrespective of the motives which lead a state to enter war.

This is not to say that a state’s motives are wholly irrelevant though. We may praise the war itself for saving lives or ending injustice, but a state which enters war for morally repugnant reasons should be condemned and perhaps even punished afterwards. This is because the state demonstrates a bad moral character, and this bad character gives us reason to expect that state to be the cause of future wrongs. Again, as before, the intentions and reasons offered indicate the moral character of the agent (in this case, the state), but this does not automatically translate to the moral assessment of acts, including war, carried out by said agent.

Therefore, right intention, as it is traditionally understood, cannot be sustained. It is indeed a good thing, or at least a better thing, for states to enter wars with good intentions, but only insofar as this makes the states themselves morally better. Wars, as actions, cannot be judged based on the internal mental states of those who wage them. Rather, wars, like all actions, must be judged good or evil based on what they bring about.

6 The Value of Intentions in War and Peace

So far, there has been much argument regarding the ways in which intentions do not matter to the ethics of war and killing. However, there are numerous ways in which they do matter, and a proper treatment must make mention of these as well.

We have already indicated throughout that intentions do matter insofar as they inform us of the character of an agent. All things considered, it is morally better to have good moral agents than to have bad ones. This would even be the case if both types of agents performed exactly the same actions, all of which were morally good or right. The reason for this is simple; more good things is better than less. If only the actions were good, but not the agents themselves, then there would be less good than if both the actions and the agents were good. For this reason alone, it is morally preferable to have morally motivated agents rather than
agents who are motivated by immoral or amoral reasons.

A second reason why intentions do (or may) matter in war, and the reason which I believe Purves et al. are ultimately getting at, is that agents who act for the right reasons perform actions which have moral worth, or at the very least, which have more moral worth.

This idea is rooted in the Kantian ideal that a morally worthy action is one which is performed from duty. What this means is roughly that the agent performing the action does so simply because it is the right thing to do, and not because the agent wants to, or will gain from it, or is afraid of punishment or social ostracism. The agent recognizes that the moral law demands a certain response, and acts accordingly, without requiring any other impetus. This is what, according to Kant, gives an action moral worth.24 I believe there is some truth to this sentiment, and to the related idea that actions done for the right reasons are morally worthier than those which are done for other, perhaps selfish, reasons.25 For example, if I bring my wife flowers to ingratiate myself with her so that I can get a back rub later, then this seems much less worthy than if I had done so simply to brighten her day. An act done for purely selfish reasons, with no regard for the wellbeing of others, does simply seem to have less moral worth than one which has altruism (or duty) at its core. In fact, some might reject this scalar notion of moral worth altogether, going so far as to say that acts of this kind are completely unworthy.

However, the fact that the selfishly motivated action has less moral worth (or even no moral worth) does not thereby imply that it is wrong. This is where Purves et al.’s argument goes wrong. They conflate the moral worth of something with its moral status, but these are distinct valuations. To see this fully, let us consider a classic example that Kant himself discusses: helping one’s friends. If I carry out some action to help my friends because they are my friends, and not because duty demands it, then this constitutes an action without moral worth. However, this does not automatically make the action wrong! What makes an action wrong, according to Kant, is that it violates the requirements of duty, not that it fails to be motivated by duty. As long as my action is in accordance with duty, it is permissible (perhaps even obligatory), regardless of my motivations. However, my motivations determine whether this permissible (obligatory) act is also a morally worthy act. Thus, the two are distinct and cannot be treated as one and the same. To do so would absurdly imply that the only morally permissible acts are those that are morally required, because those are the only actions which can theoretically be motivated by duty. Anything we do which is morally optional
is incapable of being done from duty – that’s what makes it optional – and so would by default be wrong if moral worth and moral rightness were the same. Therefore, the moral worth of an action cannot be used as an ersatz evaluation of its rightness.\textsuperscript{26}

Another, less heavily Kantian way to explain the distinction is by way of the concepts of blame and praise. Right and wrong are simply evaluative. Actions which are right are right, but we do not necessarily praise those actions. The same holds for actions which are wrong. However, actions which are done for the right reasons are deserving of praise, even if they are wrong.\textsuperscript{27} This is because, by focusing on the reasons for action, we look to the character of agents, and a good character can (perhaps ought) to be praised, even when it does evil. We may (perhaps should) attempt to explain why the wrong action was in fact wrong, but it still holds a sort of moral worth,\textsuperscript{28} and this makes it praiseworthy. The contrary holds for actions which are right but done for the wrong reasons; we should evaluate them as right, but blame the agent for being motivated by bad reasons.

To see how this would function, let’s return to the examples of noble kamikaze and sadistic marine discussed above. We argued that noble kamikaze acts wrongly, because he furthers grave injustice and violence, while the sadistic marine acts rightly, because his actions help to save many and deliver even more from tyranny and oppression. However, the actions of noble kamikaze are praiseworthy, because they are done from a motive of duty, honor, loyalty to one’s people or family, etc. Likewise, we ought to praise the noble kamikaze for no other reason than that he acts for good reasons. We also have reason to try to convince him he is acting wrongly, and failing to do that, to fight him, but we can recognize in him an opponent deserving of respect and honor, and indeed praise. This sentiment is one which has a long tradition in the history of war and conflict. Generals and soldiers alike have always differentiated between just wars and honorable or praiseworthy soldiers. What makes a soldier honorable or praiseworthy is nothing more than that he fights for the right reasons, giving his actions a sort of moral worth, and rendering him worthy of our praise.

The sadistic marine, on the other hand, acts in quite the opposite way. What he does is right, and we have good reason to encourage him to continue to do those things. However, his actions hold no moral worth because they are motivated by the evilest of sentiments. For this reason, we should disdain him and try our best to help him cultivate a stronger moral character, which is more inclined to act for the right reasons. We
shouldn’t hinder his actions which promote the just cause, but we should try to make him see that his reasons for actions are, morally speaking, very flawed.

In this way, right actions and morally worthy actions can come apart. It may be the case that those motivated by the right reasons are more likely or inclined to do the right thing (and vice versa), but it is equally possible that good motives lead to wrong actions or that bad motives lead to right ones.

7 Intentions and Practical Considerations

Before concluding it will be useful to also examine how intentions affect the practicalities of moral decisions in war. This is especially important because war is such an uncertain enterprise, and purely theoretic discussions which only deal with cases of perfect knowledge or perfect foresight are likely to lead us astray. A satisfactory account of the just war must not only tell us what an ideal observer would do or judge to be right, but must also be useful to those of us who actually have to make wartime decisions under situations of uncertainty and risk. In discussing the practical considerations associated with intentions and war, we will explore two main points; first, a further objection to the argument of Purves et al., and second, a practical argument for how intentions do matter for the moral assessment of acts in war.

As argued above, the position of Purves et al. is flawed in that it mistakes the morality of an agent’s character or the moral worth of an act with what makes an act right. These are all separate concerns which relate to different evaluations and different morally relevant things (acts, outcomes, mental states, intentions). However, the fact that we are concerned with morality in war gives rise to a practical problem with the position of Purves et al. as well.

War is a dangerous enterprise, with the possibility of death looming for nearly every soldier deployed to a combat area, as well as many of the civilians who regularly accompany fighting men (e.g. security contractors, analysts, engineers, medical professionals, etc.). No matter how much technological or material superiority a fighting force enjoys over its enemy, there is always the possibility that one be unlucky enough to be fatally wounded. Now, all other things being equal, if a certain number of men will inevitably die as a result of some military engagement and we are in a position to choose which will die, we have good moral reason to choose those men who are, as agents, least moral. Put differently, we
have moral reason to distribute harms to the immoral. To be clear, this
is not to say that immoral agents are deserving of harm or death, or that
we ought to harm or kill them. What this says is that if some must be
harmed or die, better the bad than the good.

At first glance, this likely seems a harsh sentiment, but it is one which
is well grounded in both consequential and deontological ethical theories.
On consequential grounds, the agents with flawed moral characters are
more likely to cause harms (or, more generally, bads) than those with
good characters. At the very least, a good moral agent is more likely
to follow moral rules which, regardless of their particularities, gener-
ally promote welfare, freedom, autonomy, and mutual respect. In short,
good agents promote goods of a wide variety, whereas bad agents may
not. On deontic grounds, we have reason to put bad agents in harm’s
way rather than good ones, because the good ones, in virtue of their
moral development, will almost certainly be more morally innocent and
therefore harms to them will be harder to justify. This is not to say
that all harms to a morally bad agent are by default justified, as this
would undermine much of the morality of war. However, it does mean
that the justification for harming or killing a bad agent need not be as
strong or overwhelming as that needed to justify harming or killing a
good agent, because ex hypothesi the bad agent is less innocent than
his good counterpart. The underlying deontic logic at play here is the
idea of discounting harms based on one’s liability to be harmed. Now,
both the concepts of discounting and liability are well-developed in the
literature and cannot be satisfactorily explored here, but suffice to say
that harms to a fully innocent agent will always count fully, whereas
harms to a partially non-innocent agent may be discounted (as per your
preferred deontic notion of discounting), and as such these harms are
easier to justify morally.29

Now, under the example racist soldier, upon which Purves et al.
built their argument, the racist and non-racist soldiers act in exactly
the same fashion, bringing about exactly the same results. We have
already argued above that this renders their actions exactly the same,
morally speaking. We can now add to this that, pace Purves et al., it
would actually be morally better to deploy racist soldier than it would be
to deploy his non-racist counterpart, because the deployed soldier might
die in combat. Because we know that anyone we send is under a risk of
death, we have good moral reason to send the less moral of two agents
provided they will act in the same ways. In everyday life, this may not
be true – we may always be better off picking the morally superior agent
in day-to-day affairs. However, because wars involve significant risks of harm and death, we ought to lay those risks and harms at the feet of those who are generally morally bad agents provided they will act the same as morally good agents would.

The caveat, "provided they will act the same as morally good agents", brings us to the final point, why intentions do matter morally in the assessment of actions in war. In all the examples discussed so far, we have simply assumed that the agents will act in certain ways and have certain intentions, without making any mention of how the latter may impact the former. Now, simply assuming that intentions and actions can be wholly separated and treated in a completely independent fashion is fine for a theoretical discussion, and can help illuminate subtle distinctions which are normally not apparent. However, a complete account must bring them together again, as one’s motivations for action usually do impact on the actions themselves.

It is in principle possible for a racist and non-racist soldier to both act in exactly the same way when they fight in war. However, this is highly unlikely. If a soldier viscerally hates those he is killing, as does the racist soldier from Purves et al.’s example, then it becomes difficult to believe that he really will act in exactly the same way as another soldier who does not share this hatred. We suspect, and rightly so, that the racist soldier has likely used more force than necessary, or miscalculated in his proportionality judgments (perhaps because he highly discounts the moral value of pain inflicted on so-called “lesser” people), or perhaps targeted those he should not have. In short, we expect him to violate one of the many norms of war already on the table (e.g. necessity, proportionality, discrimination) because of his hatred and the fact that he is only fighting as a result of that hatred.

This is why the intentions and reasons for action do matter in war, because they alter the courses of action one is likely to take. The racist soldier will almost certainly not act in the same ways a non-racist soldier would, and for this reason we believe we should not send him. Granted, it is natural to say (and perhaps even think) that his bad intentions are what make it wrong to send him, but when looking to his intentions we are simply using a shorthand for expressing what we expect him to do. Because he is racist, we (likely accurately) believe he will violate the in bello norms of war. This is what makes his actions wrong, not his intentions. Now, it is true that any violations of the norms of war are likely a result of his bad intentions, given that another soldier without those intentions would likely act better. However, what is (possibly)
wrong with his actions is that they are more likely to violate norms, not that he has some bad reasons for them.

This is the sense in which intentions do matter to the ethics of action. Acting based on morally bad reasons increases the likelihood of carrying out actions which are patently wrong, and this grounds the intuition that we should not deploy racist soldiers or elect racist politicians. It is not simply the case that everything the racist soldier or racist politician does will be wrong because of their racism, but we have good reason to believe that their racism will make them overlook morally salient features of certain actions. This is what grounds the belief that their actions are wrong because they are based on the wrong reasons. However, this does not say that intentions matter in principle, merely that intentions matter insofar as they affect actions and outcomes, a conclusion that is at odds with the claims of Purves et al.

8 Concluding Remarks

If at all possible, we should not deploy racist soldiers, or sadistic marines, or any agents who we know are acting for the wrong reasons. This, however, is not because their actions are by default morally wrong. Rather, agents who act for the wrong reasons are far more likely to act in ways which are morally problematic. Agents who demonstrate good moral characters, on the other hand, are likely to follow the norms of war and only fight as long and as hard as is necessary to secure the just aims. This is precisely what a theory of the just war demands, and grounds our thinking that intentions do matter. However, we must always bear in mind that the intentions only matter for actions insofar as the intentions actually affect the actions. If two agents really will do exactly the same things in war, then we ought to deploy the less morally developed of the two, given that every soldier faces serious risks of harm and death. In order for it to be plausible that intentions matter in principle, it must be shown that it is better to impose a risk of death on a virtuous agent, even though his evil counterpart would have done just as well in his place. Until such an argument is presented, we ought to judge actions in war based on outcomes alone, and only use intentions to judge the character of agents.
Acknowledgements

Much of this work was carried out while at the University of Bayreuth, and I am very grateful for their support. In particular I would like to thank Fritz Gillerke and Alice Pinheiro-Walla, whose thoughts and objections have helped me to refine many of my original ideas. I must also thank the anonymous reviewers for their helpful comments and suggestions, as well as their patience. Finally, I must thank my wife Anna, who somehow always has time to listen to my crazy opinions.

Notes

1 [14].
3 See, e.g., [5, 6, 15].
5 Right Intention is explicitly discussed in [13, 12] and [18], and finds its way into many discussions of morality and war via the doctrine of double effect. See, for example, [4] and [8].
6 A notable exception to this understanding is provided in [16]. Steinhoff persuasively argues that “a just cause is not an end to achieve, but a set of conditions that the war satisfies” (p. 39), and as such, one can at most fight with a just cause, but not for a just cause. Steinhoff’s arguments are compelling and I fully agree with his conclusions, but they represent a minority position, and so I choose to follow the standard interpretations of just cause and right intention in order to make the arguments as broadly applicable as possible. At any rate, the conclusions I reach align perfectly with Steinhoff’s.
10 Some so-called just war “traditionalists” may take issue with the claim that Rommel’s soldiers fight unjustly, as the justness of a war itself is argued to be irrelevant to the justness of particular acts in war. However, for the sake of brevity, we will ignore this debate. For the classic formulation of the traditionalist position, see [18], esp. ch. 3. For more current exposition of the traditionalist position, as well as objections to it, see [3] and [11] respectively.
12 Ibid, emphasis added.
13 Ibid.
14 Ibid.
15 I begin in this way because arguments attempting to directly show that intentions do (not) matter morally for the evaluation of action tend to be circular and
unconvincing. For example, Purves et al. simply assume that their thought-experiment, *racist soldier*, demonstrates that intentions matter in this way, when in fact it could be used to demonstrate the exact opposite, provided one has contrary intuitions. By starting with the examination of where intentions clearly do matter, it will be possible to more conclusively say where they do not matter.

16 Thanks to an anonymous reviewer for suggesting this example to me.

17 Again, just war traditionalists like Walzer may take exception to this, because of their commitment to the *moral equality of combatants* thesis, which maintains that the justness of conflict is irrelevant to the permissibility of individual soldiers’ fighting in that conflict.

18 In this way, there is also a disturbing asymmetry to the principle of right intention or acting for the right reasons, because while bad intentions/reasons supposedly taint the moral status of an otherwise good act, such that it becomes impermissible, good intentions cannot raise an otherwise bad act to the level of permissibility.

19 Some may think that the kamikaze’s failure to recognize the unjustness of the war he is fighting renders his moral character “bad”. Put differently, failure to reason well about moral facts is, according to the objection, a flaw in one’s moral character. However, this confuses epistemic flaws for moral ones. The kamikaze fights because he wants to do what is right, but he is epistemically unable to correctly judge what is in fact right (likely due to factors outside his control; e.g. propaganda). This epistemic failing does not necessarily or implicitly translate to a moral failing. At the very least, the epistemic failing cannot wholly taint his moral intentions, so his character is at least better than that of another agent without such motivations.

20 Whether we are concerned with value in terms of *quality* or *quantity*, the goods the *sadistic marine* contributes to are certainly of a high order.

21 Compact arguments along similar lines reaching similar conclusions can be found in [16] pp. 39–41 and [17] pp. 27–28.

22 Note that the traditional aim of right intention was not ensuring that a state acted justly, but rather reconciling the practice of war with Christian values and principles. “Hence Augustine says: ‘Among true worshippers of God, those wars which are waged not out of greed or cruelty, but with the object of securing peace by coercing the wicked and helping the good, are regarded as peaceful.’” ([1] p. 241) Thanks to an anonymous reviewer for reminding me of this.

23 Whatever it may mean to discuss the “internal mental state” of a state.

24 [9], 4:390 and 4:398.

25 For an insightful exploration of the value of intentions and reasons for action, as well as the distinction between acting from duty vs. acting for the right reasons, see [10].

26 For an excellent treatment of Kant’s moral philosophy, see [19]. For the explicit discussion of the differences between *from duty* and *in accordance with duty*, as well as work on how motivations and moral worth relate to these concepts, see sections 3–4 of chapter 1, pp. 26–42.

27 Note that it is crucial that we discuss actions done *for the right reasons* and not *morally worthy* actions, as moral worth is conferred only on actions which are done *from duty*, and so presupposes that the action is morally right. Therefore, it does not allow for this type of fine-grained examination.
28 Again, the qualification is necessary in virtue of the point made in the above endnote.

29 For an in-depth influential discussion of discounting and liability see [11], especially Chapter 4.

30 Again, it is worth stressing that the arguments presented are only meant to apply to cases of war and killing. Intentions may matter in principle in everyday moral decisions, as there is no, or at least less, risk for the agent acting.

Nathan G. Wood  
PhD Fellow of the Research Foundation - Flanders  
Ghent University

References


Abstract
Imagine a world where everyone is healthy, intelligent, long living and happy. Intuitively this seems wonderful albeit unrealistic. However, recent scientific breakthroughs in genetic engineering, namely CRISPR/Cas bring the question into public discourse, how the genetic enhancement of humans should be evaluated morally. In 2001, when preimplantation genetic diagnosis (PGD) and \textit{in vitro} fertilisation (IVF), enabled parents to select between multiple embryos, Julian Savulescu introduced the principle of procreative beneficence (PPB), stating that parents have the obligations to choose the child that is expected to have the best life. In this paper I argue that accepting the PPB and the consequentialist principle (CP) that two acts with the same consequences are morally on par, commits one to accepting the parental obligation of genetically enhancing one’s children.

Keywords: Human enhancement, genetic enhancement, Julian Savulescu, designer babies, procreative beneficence, CRISPR/Cas, consequentialism

1 Introduction
Imagine a world where everyone is healthy, intelligent, long living and happy. Intuitively this seems wonderful albeit unrealistic. However, recent scientific developments in genetic engineering, namely CRISPR/Cas bring the question into public discourse, how the genetic enhancement of humans should be evaluated morally. In 2001, when preimplantation genetic diagnosis (PGD) and \textit{in vitro} fertilisation (IVF), enabled parents to select between multiple embryos, Julian Savulescu introduced the principle of procreative beneficence (PPB), stating that parents have the obligations to choose the child that is expected to have the best life [11]. In this paper I argue that accepting the PPB and the consequentialist principle (CP) that two acts with the same consequences are morally
on par, commits one to accepting the parental obligation to genetically enhance their prospective children, i.e. embryos, a position Savulescu doesn’t explicitly endorse but I argue is committed too. As the argument I provide may seem like a slippery slope, the largest part of this paper is tasked with responding to a multitude of possible objections. I will argue that even though there are differences between the application of PPB and genetic enhancement of embryos, they do not demarcate a relevant moral difference between the two in respect to their obligatoriness.

This paper is structured as follows. In Section 2 I define and explicate various concepts that are related to genetic enhancement, i.e. human enhancement, wellbeing and the principle of procreative beneficence. Technological specifics and explanations of how genetic enhancements work will for the most part be omitted in this paper dealing purely with a normative question, i.e. should parents genetically enhance their children if the technology was available. In Section 3 I formulate and explain the argument that those who want to be parents have the obligation to genetically enhance their embryos. In Section 4 I anticipate and evaluate a number of possible objections including, inequality, freedom and welfare in the society. In Section 5 I conclude that even though the consequences of the PPB application and genetic enhancements in embryos might require different kinds of policies, these differences do not provide sufficient reason to give the application of PPB and the parents using genetic enhancements on their embryos a different moral standing.

2 Important concepts

According to Savulescu, Sandberg and Kahane there are two main senses in which the term enhancement is used. “Functional enhancement, the enhancement of some capacity or power (e.g. vision, intelligence, health) and human enhancement, the enhancement of a human being’s life” [10:, p. 3]. In most discussions surrounding the term enhancement, people refer to the functional definition. In all cases of the functional enhancements it is an open question whether the improvement of some capacity makes the life of the person receiving the enhancement actually better. It, therefore, is an open question what the moral stance on any particular form of enhancement is. However, proponents of enhancements usually use the latter definition. After all, Savulescu argues, it is not “disease […] which is important, but its impact on a life in ways that matter which is important. People often trade length of life
for non-health related well-being” [11; p. 419]. The concept of human enhancement, employed by Savulescu and others, is comparatively less straightforward and requires further clarification as it is crucial for my argument. Savulescu, Sandberg and Kahane jointly propose the following definition for human enhancement; “any change in the biology or psychology of a person which increases the chances of leading a good life in the relevant set of circumstances” [10; p. 7]. According to this welfarist definition, a conception of normality is not needed to define enhancement as we do not define an enhancement as being better than normal, but simply being better full stop. Medical treatments are therefore just a form of enhancements [10; p. 8]. Opponents of enhancements, on the other hand are trying to argue that enhancing is something going beyond treatment and is therefore not morally demanded. The welfarist definition of enhancement is crucial to understand why parents have the obligation to genetically enhance their children. Those who disagree because of adherence to a different account of enhancement are talking about something different. If a genetic alteration in an embryo is not expected to lead to a better life of that human, than it is simply not an enhancement.

Furthermore, it is important to explicate the principle of procreative beneficence (PPB): “Couples (or single reproducers) should select the child, of the possible children they could have, who is expected to have the best life, or at least as good a life as the others, based on the relevant available information” [11; p. 415]. Savulescu proposed this principle in a time when preimplantation genetic diagnosis (PGD) and in vitro fertilisation (IVF) were highly debated, trying to convince even non-consequentialists. However new technologies might enable us not only to screen for genes but also change them. In his consequentialist framework he himself thinks that parents have the obligation to genetically enhance their children [10; p. 16]. The question arises whether the acceptance of the PPB and the consequentialist principle (CP), that two acts with the same consequences are morally on par, itself entails the parental obligation to genetically enhance their embryos. In the following I shall argue that this slippery slope cannot be avoided.

A crucial part for this argument is to define the expectancy of the best possible life. This is a vague part of his argument, but what counts as an enhancement hinges on how welfare is defined for which there are several theories: hedonism, desire/preference-satisfaction, objective list theories, see Griffin [2] and Parfit [7]. Savulescu has not assigned himself to any particular theory, thus claiming to avoid some of their weaknesses:
“I have not committed myself to any particular substantive conception of the good life. That is a complex question as old as philosophy itself. I believe the best life is a life of objectively worthwhile activity that provides pleasure and is desired.” [12; p. 286].

Instead he focuses on all-purpose goods that would according to all of those welfare theories mentioned above count as enhancements e.g. “Memory, self-discipline, impulse control, foresight, patience, sense of humour, sunny temperament, empathy, imagination, sympathy, fairness, honesty, and so on, capacity to live peaceably and socially with others” [12; p. 284]. Surely some of this might harm us e.g. post-traumatic disorders in conjunction with better memory and imagination, but what matters is that they are expected to make a life go better. A final definition of welfare is not necessary for our account as long as we are able to evaluate these all-purpose goods. Even if it turns out that there is only one such all-purpose good satisfying the major theories of wellbeing, i.e. hedonism, desire/preference-satisfaction and objective list theories, the argument provided in the next section will require parents to genetically enhance it in their children. Nevertheless, as we follow the welfarist definition of what an enhancement is with medical treatments being a subset of enhancements, the set of all-purpose goods might be rather large, e.g. a functioning heart, brain, lung, etc.

3 Obligation to genetically enhance embryos

There are at least two ways to argue for the position that the PPB entails the obligation to genetically enhance embryos. One could argue that selection out of possible children entails the selection out of possible genetically enhanced embryos. After all one could possibly have a genetically enhanced child. In that case genetic enhancements would simply fall under the PPB and hence be obligatory. However, this is debatable and is not necessary to accept my claim. For my argument to proceed I will require an additional premise, which is the consequentialist principle (CP). Consequentialists are only concerned with outcomes, how these outcomes are achieved is morally irrelevant. Killing and letting die are thus ceteris paribus morally equivalent. In order for my argument to hold, I need to show that the outcomes of the PPB and genetic enhancements in embryos are morally on par. What is the crucial outcome of PPB? It is the expectancy of the best life, which we are expected to achieve with Savulescu’s all-purpose goods. To formulate my argument:
(1) Those who want to be parents have the obligation to select the child that is expected to have the best life, or at least as good a life as the others, based on the relevant available information (= the PPB).

(2) Application of the PPB leads to the creation of children that are expected to have the best life or a life at least as good as without the application of the PPB, based on the relevant available information.

(3) If act A and act B have the same expected consequences, then they are morally on par. (= the CP)

(4) Genetic enhancements in embryos lead to the creation of children that are expected to have the best life or a life at least as good as without genetic enhancements, based on the relevant available information.

(5) Therefore, those who want to be parents have the obligation to genetically enhance their embryos.

From premise (1) follows, that (2) is obligatory for those who want to be parents. As (2) and (4) have the same consequences, it follows from the CP (3) that (4) is also obligatory for those who want to be parents (5). As stated in the beginning, we are accepting the PPB, therefore premise (1) & (2) and the CP (3). If someone, e.g. Savulescu would want to refute the conclusion (5) without dropping one’s commitment to (1), (2) & (3), only two ways remain. First, denying the truth of premise (4). Second, denying that the conclusion follows from the premises as the consequences of premise (2) and (4) differ in a morally relevant way. Let me now turn to these considerations in section 4.

4 Objections and Defense

Keeping in mind that every position that refutes the conclusion, but is inconsistent with premise (1), (2) or (3), is not a valid objection to the argument I bring forward. One cannot deny the conclusion of an argument that takes the form, “if you accept these premises, you have to accept this conclusion” by denying the truth of one of the premises. The argument provided in this paper aims for just this sort of conclusion: Philosophers accepting the CP and the PPB do not have a valid escape from being charged with the slippery slope of advocating genetic enhancements. To narrow the scope of this paper I am not taking a stance
on the matter of whether the CP or the PPB is right. Such an endeavor would require more than a single volume book. I only argue that philosophers and researchers like Julian Savulescu, Anders Sandberg and Guy Kahane should either bite the bullet of advocating human enhancements or indeed stop being consequentialists advocating the PPB. The objections that follow are then nothing more but misguided attempts of stopping the ‘slippery slope’.

Let me begin with the seemingly, obvious objection that premise 4 is false. An objection often put forward is that genetic enhancements have risks that are too high or unknown and would make the life of the person targeted go worse. I argue that this hinges on different conceptions of what counts as an enhancement. First, it is clear that genetic alterations do not necessarily increase the expected welfare of an agent. However I am not concerned with genetic alterations per se, but with genetic enhancements. No serious advocate of enhancements in humans would deny the difficulties in achieving successful genetic alterations. In fact, the terminology of philosophers and researchers advocating the PPB, i.e. Savulescu, Sandberg & Kahane doesn’t allow for this objection. As noted in section 2 they employ the welfarist definition of enhancement: A genetic enhancement by definition increases the expected welfare of children. The question whether genetic enhancements thus really increase welfare is irrelevant, because premise (4) is in fact a tautology. Though this may seem like a cheap way of justifying genetic enhancements in humans, this is what proponents of human enhancements are advocating: Using genetic alterations when they are expected to improve the life of a human embryo, including perhaps open-mindedness for the future of the technology.

It can of course still be asked whether genetic alterations are technologically possible in such a way that they lead to the creation of a child with higher expected well-being. This is an empirical question and even if they currently are not, this doesn’t free us from the task of answering this question. After all many questions in ethics are formulated like this: one might argue that current technological limitations might give society an obligation to invest heavily into research as the potential benefits are enormous. However this is of no concern for my argument, because even those who disagree with genetic enhancements ever being technologically possible have to accept alleged empirical evidence. In 2015 Chinese researches released an article, showing that the gene editing of human embryos is in fact technologically possible. They used a technique called CRISPR/Cas9 which makes it impossible to tell which
genes are edited [4; p. 363-372].

Realizing that premise (4) being a tautology, cannot be false, and conditionally accepting premise (1) – (3), any further objection against the conclusion following from the premises might seem pointless. However one way to refuse the conclusion remains, that is denying that the consequences of PPB application and genetic enhancements of embryos are the same. Granted, it is obviously true that two acts may never have the ‘same’ consequences, if we consider every possible consequence. However these differences have to justify a different moral standpoint, meaning that the PPB is obligatory, while genetic enhancements of embryos are not. For two acts can be different in their goodness while, nevertheless, still both being obligatory. A familiar example in line with PPB would be the obligation to send your children to school and the obligation to feed them.

Objections of this sort can be categorized in two ways. First, the expected welfare of the child might differ between selecting and enhancing an embryo in a morally relevant way. Second, they might only be morally equal in one respect, i.e. expected welfare of the child, while leaving out other relevant consequences that differ between the two and justify a different moral standing. Some objections might overlap as consequences such as freedom or equality and their implications on the welfare of the targeted embryo and overall/average welfare. The remainder of this paper has the purpose to explore and debunk a multitude of such objections.

4.1 Degree and Precision

Concerning differences in wellbeing, Parens objects that genetic alterations might be morally different from other welfare increasing procedures of offspring because they have the potential to be much larger in degree and precision [6; p. S7]. However concerning premise (4) this objection is unsuccessful, because the consequence we want is explicity the life that is expected to be best. If genetic enhancements do so much better than mere selection than Parens’ objection would rather support a moral priority of genetic enhancements against mere selection. However some might object that genetic enhancements are less reliable in their effects, therefore unprecise. That however is an empirical question and doesn’t change the fact that these genetic alterations are expected to make the targets life better. Though this is the most obvious objection and should therefore be tackled first, it is also the least harmful to the
argument provided in this paper.

4.2 Risk

Another objection often put forward against genetic enhancements is the risk objection, i.e. that the PPB has far less risks to its application compared to genetic alterations. Of course genetic alterations currently bear risks much greater than mere embryo selection by IVF and PGD, which do not change the genetic constitution of the embryo. However when genetic alterations are expected to be overall more harmful than beneficial, they do not count as genetic enhancements and are therefore ruled out. If we cannot determine probabilities, than we cannot judge the expectancy of the best life anyway. This critic of human enhancements is rather meaningless against the welfarist definition of enhancements. When it comes to risk, the opponents of genetic enhancements are attacking a straw-man position that proponents of genetic enhancements do not adhere to. Risk, degree and precision are empirical concerns not only for the opponent of genetic enhancements, but also for enhancement advocates who DO take them serious.

But even if the risk of harm is currently unknown or outweighs the potential benefits, this doesn’t indicate that genetic enhancements should be avoided. One could rather argue again that we should do research until the expected benefits outweigh the medical risks of harm not stop research altogether. Also the potential benefit to be gained by genetic enhancements is much larger in degree than mere selection. Demanding a risk of zero to the targeted agents is over-demanding, otherwise medical treatments would never be at the point they are today. Still, this is a factor that affects the expected well-being negatively and therefore has to be considered in judging whether a genetic alteration counts as an enhancement.

One might further object that, contrary to somatic gene therapy, which only affects the targeted individual, germ-line interventions will affect the offspring of the targeted embryo and the offspring of his offspring and so forth. . . However we are enhancing all-purpose goods like intelligence and health. It seems implausible to claim that all-purpose goods are always expected to increase welfare, but might not do so for future generations. Even so, if in the future new scientific advances change our subjective probability assessments, genetic alterations that did not count as enhancements, could do so in the future and vice-versa. However, it would then be the obligation of our hypothetical parent’s
off-spring to ensure, that their own off-spring has the best life possible and perhaps undo some of the changes to their gen-code. The PPB does not indicate an obligation towards grandchildren, but only to our own children. Different times and cultures may very well render different genes beneficial, think of intelligence and the muscular labor in former centuries.

Maybe it will be the case that those enhanced will be less well off than unenhanced humans would have been, because they see themselves as objects rather than subjects. A “natural” patient with paraplegia may be better off than a “super-enhanced” guy with superiority in health, intelligence, strength etc. but cannot deal with the fact that he is created artificially, perhaps becomes an alcoholic or worse. If we had reason to give this outcome a certain probability, it would have to be calculated in whether we can judge a genetic alteration to be an enhancement or not. Therefore, higher risks are no reason to justify a different moral standing from PPB.

4.3 Harm

However even though the expected welfare of genetic enhancement in embryos and the application of PPB are the same, the issue of potential harm to the embryo could still justify a moral difference between the two. The PPB seems to capture the intuition of parents being obligated towards their children’s welfare. Harming would indeed be quite contrary to this. Savulescu argues that genetic enhancements can harm an embryo in a way mere selection cannot. He asks us to imagine an embryo A, who was selected for existence but later develops cancer. As long as its life is worth living, it cannot be said that we have harmed him, for he would not have existed otherwise. However this still seems to be morally wrong. If we genetically alter an embryo B and he develops cancer (assuming his life is worse than it would have been without genetic alteration) then we would have harmed him [11; p. 422]. Unlike the objection to premise (4) we cannot disregard the harm objection by hinting at the welfarist definition of enhancement. Genetic enhancements are expected to lead to the best possible life. However harm is concerned with the actual life, not the expected welfare. In this respect PPB and genetic enhancements differ because the harm objection is successful against genetic enhancement while it is not against PPB.

However Savulescuses argument also implies, that parents that selected an embryo with the expectancy of the highest welfare and turns out to
have a higher wellbeing than the others, cannot claim that they benefitted him. On the contrary parents that arranged their embryo to be genetically enhanced and the embryo turns out to benefit from it, can take credit for this. Imagine a couple both bearing a defective gene that would certainly cause blindness in their child. IVF and PGD would not be able to help them (and the child), however a genetic enhancement would. According to the CP, failing to select for sight and blinding a child would be morally equal, i.e. impermissible as they share the same consequence. If such enhancement technology were available children could justifiably condemn their parents, for having harmed them in the sense that they did not give them this benefit. This might seem odd at first, but using an ordinary example, we condemn parents who do not send their children to school in order to ensure their education. In this case we also use the term harm. The claim then that there is a moral distinction between harming and failing to benefit requires a special kind of argument that may not be compatible with the CP. I take on the view that there is no relevant moral difference. As such the harm objection that aims to establish a relevant moral difference between genetic enhancements and applying the PPB raises a benefit objection against the PPB that parents do not actually benefit their children.

However, I have an additional argument against the harm objection and that is the irrelevance of harm and benefit altogether. As indicated in section 3 the morally relevant part of the PPB is the expectancy of the best life, I argue that it is not selection of the best expected life that is morally relevant but the creation of the best expected life. Using again the example of a blind child or the child not being sent to school, it is obvious how the term benefit and harm seem to be used interchangeably. If refusing to provide benefits is the same as harming, than these categories are not morally exclusive. Therefore if we take premise (3) serious, we cannot argue for a moral difference between selecting and altering. The only thing that matters is the creation of the indicated consequence welfare. Those who disagree would have to solve the non-identity problem of Parfit, which suggested that harming future individuals is not possible, because in acting in that way the people existing in the future are completely different from those who would have otherwise existed, thus would not have existed otherwise. This is under the assumption that small events can change history dramatically. The same can be said about long-term investments that would only benefit generations far in the future. However, even when we do not harm future generations, we still value actions as wrong that lead to worse lives than otherwise could
have existed [8; p. 100-115]. The analogy to the PPB and genetic enhancements is obvious, as no one is harmed that would have otherwise existed. Therefore I conclude that what matters for parents is not a life being beneficial or harmful for their child but the creation of a “best life” in a welfarist sense – providing further support for the thesis that there is no relevant difference in their moral consequences. Let us now turn to possible moral differences between selecting, i.e. the application of PPB, and enhancing an embryo in areas other than the welfare of the embryo. In the following I will explore freedom, equality and the welfare of society.

4.4 Freedom

Let us first consider consequences like freedom, autonomy or missing consent. The majority of philosophers refuse the view that it is impermissible to have children on grounds of their lacking consent for being born. The notion of requiring consent from an embryo appears to most philosophers unnecessary or perhaps more so incoherent. However in the case of genetic enhancements on embryos this matter might be different, for they could have existed without genetic enhancements. Still, there are many therapies/treatments in medicine, that are often deemed to be obligatory, even though there is no consent e.g. vaccination. Also everyone should be familiar with parents sending their children to school, even though their children might not have given them their consent. As a child I certainly wondered who gave my parents my permission to so. But even granting that we can meaningfully speak of consent here and it being morally relevant, the consequentialist framework requires us to weigh the negative consequences against benefits of such genetic alterations in order to justify a violation of consent. In fact freedom to choose can be viewed as instrumental or even constitutive to welfare, but this objection would be ruled out by the welfarist definition of enhancement. If a genetic alteration undermines such it cannot count as an enhancement. The question then becomes whether a genetic alteration does so in general. Thinking of deleting a gene that causes deafness or blindness suggests otherwise. In medicine there is often a clash between respecting the autonomy of the patient and his welfare. But as Bostrom and Roache highlight, considerations like missing consent cannot be applied for embryos that by definition cannot give consent. Instead they argue, we should make decisions in a way that would be in their interest, so thinking in terms of their welfare [1; p. 22). Habermas claims that this
goes against the freedom of the embryo [3; p. 62]. He calls this denying the opportunity of being “the undivided author of his own life” [3; p. 63]. Again Bostrom and Roache argue that enhancements would not decrease autonomy, but rather have the potential to increase it [1; p. 21]. Just as education makes us more autonomous, genetically increased intelligence would serve the same purpose. If the mere existence of genetic enhancements undermines our feeling of being autonomous beings, than this is not an argument against genetic enhancements, but rather exposes our concept of autonomy as nothing more than an illusion. If children feel like they have to fulfill the plans of their parents, than this might limit them in their freedom to be the “undivided author” of their own lives, but this is an objection to a particular treatment of children. Genetic enhancements have the aim to increase rather than limit ones possible choices in life. If genetic enhancements make us free to do things we would otherwise not have been able to, then this seems to be even further support for genetic enhancements.

4.5 Equality

Suppose that besides the PPB, equality is also a morally relevant consequence. There is a rather famous critic against genetic enhancements from Mehlman and Botkin, i.e. them being too expensive and therefore even when all parents follow their obligation, the children of rich parents will be far more enhanced or those of poor parents won’t get any enhancements at all (1998). Why could this objection be applicable to genetic enhancement but fail against the PPB? Applied as an objection against PPB, Savulescu explains that one would have to argue that because selecting the child with the expectancy for the best life will lead to more inequality, parents would have to create a child with worse life prospects [12; p. 288], which seems to be counterintuitive. If everyone applies the PPB there might rather be more equality rather than less by eliminating the natural lottery and additionally making everyone better off. Under this argument we could accept both equality and the PPB as compatible. However with genetic enhancements this might not be possible. First let me defend my argument against the intrinsic value of equality. Equality unlike freedom or welfare as a concept always hinges on the relations between individuals. Contrary to this, an enhancement that makes someone less equal, say by bringing a genetically altered embryo into existence with the disposition for intelligence far above average (other things being equal), would increase
inequality in genetic makeup. Concerning the worry of unequal access through wealth, Savulescu argues that this is not an objection to the PPB hence genetic enhancements, but to all purchases of benefits like better education or healthcare. If equality matters then these benefits should be available to all [12; p. 288]. The same obviously applies to genetic enhancements. This is neither an objection against the PPB nor the parental obligation to genetically enhance their embryos, but to how equal the access to technologies in society is. Both consequentialist views can very well coexist, e.g. if the state ensures equal access while parents are only concerned with their child. Furthermore there are reasons to believe that genetic enhancements enhancement will not affect inequality at all, or even decrease it as Bostrom and Roache [1; p.16] suggest by making people “more equal”, like it is the case with Modafinil [9], the same could hold for genetic enhancements. How inequality could affect welfare, will be adressed in the following section.

4.6 Welfare

Another consequence of applying PPB is the effect on overall/average welfare. The objection against the PPB is that it is too much focused on the individual as opposed to the lives of everyone [12; p. 287]. Of course the same objection can be applied to genetic enhancements. In fact even if we state genetic enhancements might be good for the one who gets them (by definition), they could decrease the overall/average welfare and vice versa. Now referring back to freedom and equality we can account for their instrumental or constitutive value to welfare in this part. The implication of both equality and freedom on the individual welfare is by definition accounted in what counts as an enhancement. However both could have morally relevant implications on the overall/average welfare.

For instance, such a moral principle might lead to discrimination of the unenhanced. However, Savulescu response is that discrimination does not show that the PPB is wrong, but rather how people treat each other, which is a different topic. Genetic enhancements in empathy, sympathy and other “moral” capacities could very well decrease discrimination to a level that is even lower than the discrimination we face in an ‘unenhanced world’ [12; p. 288]. Also, the optimistic outlook that increased intelligence, including emotional intelligence, should make ‘the enhanced’ less likely to discriminate should not be disregarded. Even if, in virtue of introducing genetic alterations as a new technology, this leads to a new form of discrimination (against the unenhanced), we might sig-
nificantly decrease other forms of discrimination once ‘moral’ enhancements are introduced. Though, akin to parents denying their children vaccines, I suspect that it is much more likely that if there is any new form of discrimination, then it will be directed against parents denying their children the possibility of a better life, i.e. moral condemnation. In analogy, I argue that just as we should not let scientific research into vaccines be influenced by anti-vaccines parents, fearing moral condemnation, the introduction of genetic alterations for humans should not be stopped by set of parents unwilling to use the new technology. Both act against the interest of their child and perhaps just as in the case of vaccines deserve moral condemnation. In fact, vaccines are just a form of enhancement.

The ‘discrimination-objection’ further requires the possibility of distinguishing between the enhanced and unenhanced. With the new technology CRISPR/Cas9 it is, however, impossible to find out which genes were altered or edited [4]. The unenhanced might not even know that they have not received enhancements. Even if there won’t be blatantly obvious discrimination of the unenhanced, we may argue that there will be discrimination in competition like work, because the unenhanced might on average simply be less efficient. The discrimination then would be based however, on differences in skill rather than a naïve enhanced/unenhanced distinction. But first such a treatment does already take place and is generally not regarded as discrimination. If some parents are able to afford a better or further education for their children, e.g. tutoring, we neither prohibit such practices nor are we able to provide good counter-arguments for parents who insist that they are obligated to provide their children with the means to get ahead. Most political philosophers take it as the task of the state to make education accessible to those who would otherwise not be able to afford it, including perhaps tutoring. Rather than denying the parental obligation of enhancing one’s children, the ‘discrimination-objection’ provides a very good argument for an egalitarian policy making genetic enhancements available to all.

Savulescu accepts that some enhancements might make the individual better off, but harm the rest, e.g. by making someone manipulative or cunning [12; p. 287]. For example parents might choose to alter the DNA of their child in a way that makes them more cunning. Resulting in “him” being better off, while the overall/average welfare decreases. However, that argument can simply be taken in the opposite direction. We might want to alter the DNA of embryos that will expectedly lead
them to have worse lives, but will benefit the overall/average welfare. This of course would also be an issue for the PPB. One cannot hold this view and simultaneously be in favor of the PPB. Though the argument might successfully undermine the PPB, a defense of said principle exceeds the scope of this paper. For a defense of the PPB I suggest the work of Savulescu [10, 11, 12]. Even so, this might be an issue for the state that should enforce restrictions, while parents are primarily concerned with the welfare of their own child. The effects of one particular enhancement on the overall/average welfare of society must in the eyes of parents seem negligible. These arguments are usually stated in a much broader sense, i.e. the availability of the technology itself. In fact, even with restrictions on the technology the underlying parental obligation doesn’t cease to exist. If the state is justified or obligated to stop such genetic enhancements as a whole, be it by appealing to equality, freedom or overall/average welfare, it is an open question whether that actually frees parents from their obligation to genetically enhance their embryos. There are several reasons to think that a prohibition of genetic enhancements by the state doesn’t free parents from their obligation. First, it would be over demanding to ask from parents to act against the best interest of their child on grounds of marginal effects on the overall/average welfare in the society. Second, it would be an incompatible view with the PPB, as parents specifically have to aim for the best possible life. Third, even when means to genetically enhance are not available, this doesn’t mean that in case the genetic enhancements were available, there would be no obligation to apply them.

5 Conclusion

To conclude, objections against genetic enhancements not weighing their costs against their benefits must fail against the backdrop of the CP. Of course maybe the PPB, thus premise (1) or the CP, thus premise (3) is false, but the aim of this paper is to establish that the parental obligation to genetically enhance their children logically follows once the PPB and the CP are accepted. The ‘slippery slope’ is genuine. Objections trying to establish that the individual costs outweigh the benefits of the targeted agent, seemed to be more successful. However, a genetic alteration that is not expected to lead to the creation of a being with higher welfare cannot be called an enhancement in the welfarist sense. Merely showing than that an enhancement in the sense of increasing some capacity is contrary to the welfare of the child does not undermine the argument
provided in this paper as it precisely rests on the welfarist definition of enhancement. However, these cases might provide us with a valid argument against exceedingly strong optimism in discovering *all-purpose goods*. Having explored the potential differences in the PPB and the parental obligation to genetically enhance their children in section 4, I conclude that even though they do not share exactly the same consequences, the differences between them do not justify a different moral standing, i.e. their obligatoriness. With the additional premise of lacking relevant moral differences between the consequences of the two procedures, we have successfully defended the conclusion (5). That is to say, those who accept the PPB and the CP must bite the bullet and accept that parents have the obligation to genetically enhance their embryos. As a consequentialist Julian Savulescu should have no problem of doing so.

**Acknowledgements**

I would like to thank Roberto Fumagalli and three anonymous referees for helpful comments over the course of the development of this paper. Furthermore, I would like to thank the students, from the University of Bayreuth and the London School of Economics, with whom I had lively discussions.

_Walter Veit_  
*University of Bayreuth*  
*Universitätsstr. 30*  
*95447 Bayreuth, Germany*  
<wrwveit@gmail.com>
References


According to Theodosius Dobzhansky’s famous dictum, “nothing in biology makes sense except in the light of evolution” [3]. On the other hand, philosophers like Ludwig Wittgenstein used to be rather skeptical concerning the relevance of evolutionary thinking to philosophy: “The Darwinian theory has no more to do with philosophy than has any other hypothesis of natural science” [6, 4.1122].

In the last decades, however, – in particular since Richard Dawkins coined the term ‘meme’ for the cultural counterpart of the gene [2] – the application of evolutionary principles has been successfully pursued in areas other than biology. The central principles are reproduction, variation, and selection [4]. Based on these principles, several models, methods, and theories of a wide range of phenomena have been developed – not confined to the realm of biology. More generally, philosophy of science, social sciences, psychology, economics, and many other areas show a growing interest in a generalized theory of evolution.

Like in many disciplines and areas of research, the publication of an introductory textbook indicates a certain stage of maturity. Such a textbook has been provided (in German) by Gerhard Schurz [5]. An updated English version is currently in preparation.

Being one of the first big conferences in this area, “The Generalized Theory of Evolution” brought together international researchers, scholars, and an interested audience to discuss the current state and trends of the interdisciplinary field of a generalized theory of evolution.

1 Facts and Figures

“The Generalized Theory of Evolution” was organized by members of the Duesseldorf Center for Logic and Philosophy of Science (DCLPS): Karim Baraghith, Christian J. Feldbacher-Escamilla, Corina Strößner, and Gerhard Schurz. The event took place from January 31 to February 3, 2018 at the House of the University in Düsseldorf, hosting approx. 100 international and interdisciplinary participants from more than 15 countries.
The conference featured seven keynote lectures (briefly described in section 2) and 34 contributed talks (see section 3). The percentage of women of the conference’s participants was approx. 20%.

Definitely a highlight of the conference was a round table discussion session: “Cultural Evolution or Social Darwinism? Prospects and Problems of a Generalized Evolutionary Theory” at the beginning of the conference. Some of the keynote speakers (see below) participated in a round table discussion on the first evening of the conference – an excellent opportunity to get to know each other, to find differences and commonalities, setting the stage for the rest of the conference.

“The Generalized Theory of Evolution” was sponsored and supported by the DCLPS, the Heinrich Heine University Düsseldorf, and the DFG (Deutsche Forschungsgemeinschaft).

More details and a comprehensive repository are available at the conference website: http://dclps.phil.hhu.de/genevo.

2 Keynote Lectures

Daniel C. Dennett (Tufts University) started off the conference with his keynote talk “Tools Making Tools: the recursive de-Darwinization of human culture”. As one of the key figures in developing and promoting the idea of cultural evolution, he traced the history of the meme concept and defended it. He also stressed that competence and comprehension need to be kept apart.

On the second day of the conference, Alex Mesoudi (University of Exeter) continued with a lecture entitled “The Viability of a Theory and Science of Cultural Evolution”. He investigated the parallels and differences between biological and cultural evolution. In particular, he argued for the feasibility and adequacy of evolutionary cultural explanations.

Gerhard Schurz (University of Düsseldorf) summarized the central pieces of a generalized theory of evolution in his lecture “Generalizing Evolution Theory: Evolution in nature and culture”. After outlining the abstract principles of a generalized theory of evolution and differences between biological and cultural evolution, he distinguished different kinds of evolutionary dynamics, depending on the selection parameters.

Brian Skyrms (University of California, Irvine) gave an overview of “Some Evolutionary Dynamics of Signaling Games”. He presented a tour de force of evolutionary dynamics and learning dynamics in different kinds of signaling games. The robustness of the results of some dynamical models calls for their applicability in real-world settings.
Next, Ruth Mace (University College London) provided fascinating insights into the field of anthropology in her talk “No Need for an Upgrade: Using the toolkit from behavioural ecology to study cultural evolution”. By looking at examples of evolutionary patterns of residence and kinship from China and Africa, she argued for the relevance of Tinbergen’s framework for studying human cultural behavior.

Thomas Reydon (University of Hannover) delivered his lecture via Skype: “Towards Applicability Criteria for Generalized Evolutionary Theories: The concept of real populations”. He raised the question as to what the requirements are for a process to fall into the domain of a generalized theory of evolution. His answer lies in a generalized concept of populations which serves as a criterion for deciding whether a process meets these requirements.

Finally, Eva Jablonka (Tel Aviv University) concluded the conference with her lecture “Culture: An Evolutionary-Developmental Approach”. She pointed out the importance of George Price’s distinction between Darwinian selection and sample selection, discussing examples from epigenetic variations and reinforcement learning, respectively. Within this framework, she elaborated a developmental systems approach to cultural evolution.

3 Contributed Talks

In addition to the keynote lectures, which took place each morning and evening, a total number of 35 talks were given by researchers with various backgrounds. This diversity was structured into slots in two parallel sessions, each with a length of one hour and a half to two hours. For each talk, 20 minutes speaking time and 10 minutes discussion time had been reserved.

A variety of interesting contributions were presented, including theoretical and experimental work from biology, philosophy, game theory, economics, history, anthropology, sociology, technical studies, and many others. The organizers managed to accommodate this diversity of topics in sessions, ranging from “Complexity”, “Communication & Language”, and “EvoDevo” to “Generalized Evolutionary Modelling”. For more details on individual talks, see the abstract book, which can be downloaded from the conference website.
4 Summary

While the theory of evolution continues to be a central topic in philosophy of biology, the generalized theory of evolution is an interdisciplinary field in which the theory of evolution itself becomes a powerful framework, within which a variety of questions from biology, philosophy, social sciences, economics, anthropology, etc. can be addressed. The conference in Düsseldorf showed how active and fruitful this area of research currently is. The scientific community is looking forward to seeing more events like this and to discussing more fascinating insights in the future. Within the generalized theory of evolution – to adapt Charles Darwin [1, p. 490] – wonderful ideas “have been, and are being, evolved”.

Notes

1 Thanks to Christian J. Feldbacher-Escamilla and Karim Baraghith for providing me with the numbers.

2 Videos of the keynote lectures and the round table, together with an introduction video, as well as some of the presentation slides can be accessed at the conference website’s repository: http://dclps.phil.hhu.de/genevo/repository.

3 Available at: http://dclps.phil.hhu.de/genevo/programme.

4 Quoted from Darwin Online, accessed 02-14-2018: http://darwin-online.org.uk/content/frameset?itemID=F373&viewtype=text&pageseq=508.

Gregor P. Greslehner
Department of Philosophy, University of Salzburg
Franziskanergasse 1, 5020 Salzburg, Austria
<gregor.greslehner@gmail.com>

References


