Once more into the breech, dear friends, once more;
Or close the wall up with our English dead.
In peace there’s nothing so becomes a man
As modest stillness and humility;
But when the blast of war blows in our ears,
Then imitate the action of the tiger;
Stiffen the sinews, summon up the blood,
Disguise fair nature with hard-favour’d rage

Harvest Moon
An Undergraduate Journal of Philosophy

University of California Berkeley
May 2003 ~ Volume 2
This work was made possible by generous grants from the UC Berkeley Philosophy Department, the UC Berkeley Classics Department, The ASUC, and the Chancellor’s Committee on Student Publications. Our deepest thanks for for their support.

We would also like to thank Tony Bezsylko, Professor Boyd, Professor Code, Ashley Daley, Julien Deonna, Jon Fong, Jessica Gelber, Richard Gilles, Mike Heath, Forrest Hartman, Linda Hicks, Chaim and Zipora Jacob, Ron Kasl, Boae Kim, Professor Knapp, Traci Knox, Berislav Marusic, Gustavo Mata, Omar Mirza, Sebastian Mikowicz, Hal Reynolds, Jessica Samuels, Danille Santorio, Aaron and Susi Sassoon, Louis Savar, Professor Scheffler, Professor Searle, Charles Taylor, Carla Yumatle, and Joel Yurdin.

Harvest Moon is not an official publication of the Associated Students of the University of California. The contents and opinions herein are those of the writers (those whose names appear before their papers) and not of the ASUC.
This journal, unique of its kind, not only provides a forum for the public dissemination of some of the outstanding accomplishments of our students, but also in the short time since its inception has had a valuable impact on our department’s contribution to the educational mission of the university. Philosophy at its best is an interactive enterprise, characterized by ongoing critical discussion and the resultant replies, clarifications, rejoinders, revisions, and new directions of inquiry. This kind of collaborative dialogue, already an important part of the intellectual life of our undergraduates, has been further enhanced and promoted by the editors of Harvest Moon in their attempts to solicit, referee and edit fresh and challenging essays from within Berkeley’s undergraduate community. In this process students learn more about the accomplishments of their peers, give them feedback and advice on their philosophical writing, and participate more fully in the activities of the department. Instructors encourage students to submit their best essays to the journal, and this in turn gives them an incentive for pursuing their philosophical interests and receiving advice and feedback outside the context of formal course requirements.

There is a new feature of Harvest Moon that exemplifies well the virtues of the kind of collegial and cooperative atmosphere that the journal itself has helped foster among our students. This is the Editors’ ‘Round Robin’ discussion. The topic this time is Nozick and Political Philosophy, and contains discussions, attacks and defenses from various perspectives of Nozick’s entitlement theory of justice. However, in one way or another every single essay in this volume also represents the results of collaborative dialogue and interaction among undergraduates, graduate students and faculty. One cannot help but be impressed by the range of topics, the richness of the discussions and the sheer philosophical interest of the articles that have been selected for publication. Without being confined to narrow disciplinary labels, these essays range over central topics in philosophical logic, philosophy of science, philosophy of mind, philosophy of language, metaphysics, political philosophy, issues pertaining to affirmative action, as well as a central issue in the history of philosophy. The topics themselves are approached in many ways, drawing upon techniques and methods from both Analytic and Continental Philosophy, and doing so in a manner that highlights just how much there is to be gained by thoughtful engagement with the legacies of both traditions. The editors are to be warmly thanked and congratulated for their exemplary contribution to the academic life of the department, and the prospective reader is urged to take advantage of the
fruits of their efforts.

Alan Code
Chair, Department of Philosophy
University of California at Berkeley
May 5, 2003
Proud, frantic, and somewhat weather-beaten, our small journal ca-
reeneed into existence last spring. At that time, the idea was simply to collect 
and present the philosophical efforts of undergraduates here at Berkeley, in 
what we hoped was a more attractive form. The process was unduly long, 
and left much wanting.

During our time here at Berkeley, we have been profoundly impressed 
and influenced by the hard work, dedication, and commitment to genuine 
intellectual development that students here have. Our respect for their efforts 
drives our effort here to improve upon last years volume. What struck us 
a year ago, as it does now, about other such undergraduate productions in 
philosophy, is the displaced sense of priorities. We simply cannot understand 
how or why students might be compelled to put so much time and work 
into conducting interviews, mimicking professional journals, or simply 
producing a journal for professional philosophers, when there are so many 
students around them they can be enriched by, and perhaps enrich. In that 
respect, we hope we are not alone; however, in regards to the form of such 
a publication, this journal does differ from other such journals. To the best 
of our knowledge, it remains the only one in the nation where students in-
dependently organize to entirely produce and distribute work in philosophy 
for, by, and of undergraduates.

So we began again, as promised, in late February. Acutely aware of 
the distance the first issue was from memory, the limited span of time we 
had to work with, and that this all might be sophomoric anyway, we have 
held meetings, found the necessary funding, refereed all the excellent sub-
missions, commented on the papers, in some cases repeatedly commented 
on the papers, been as merciful as we could to the desperate cries of “more 
time,” “more time!” and laid out the entirety of these pages. With no prior 
acquaintance with layout or design, a fifteen minute primer on the program, 
and seven sleepless nights later, we produced a more graceful product with 
a finer attention to detail. To quote Noam, “we got into the words.” That 
was one change from last year, another was the marked increase in submis-
sions. We thank all those students who submitted papers to the journal, the 
sheer volume of quality work reaffirmed our conviction, when we started 
this journal, that undergraduates are regularly creating things that should not 
be lost. The other notable change was our experiment with a “round robin” 
discussion on political philosophy, which we hope brings some rigor and 
forbearance to the central questions.

As much as this journal has, and we hope will continue to have, a 
definite purpose and aesthetic, its appeal for us has been its ethos and sense
of community. This journal is committed to the philosophical growth of the undergraduates here at Berkeley and the exchange of ideas between all its constituent members. Anyone serious about these commitments has a welcome place with this project; those who will hammer away day and night, do the work necessary, and perform the tasks needed to complete this project have control over it. We hope this sense of purpose shall endure. We are pleased to pass these pleasures, duties, and I.O.U.’s to Noam Jacob and David Cohn. They give us hope that this sort of experiment might just succeed, and they understand why we care.

First and foremost, we would like to extend our most heartfelt gratitude to our philosophy department: they have given us invaluable tools with which to pursue the things we care about, and in specific regard to this project, the freedom to stand or fall by our own way of doing so. Our deepest appreciation Professor Code for his constant support and guidance, he has helped us clarify and further our efforts in so many tangible ways. Our thanks also to Professor Scheffler for his patience, careful explanation, and good-natured humor during our frequent and oft belabored visits. Richard Gilles’ patience knows no bounds, and his thoughtful help and advice have expedited us through an enormously complicated layout process. Our thanks to Lewis Savar, for coming over at midnight to help format Marcy Milkowski’s cover art. Linda Hicks is simply amazing, we thank her for all of her kind and efficient help and advice, especially as the time has drawn thin. Jessica Samuels’ advice and support, as well as her thorough and incisive work in copy editing has proved invaluable, we hope we made it worth her time. Also, certainly, thanks for Triple Rock.

We would like to think we have created a forum in which students interested in philosophy can pursue and publish their best work. We would like to believe we have presented an alternative space in which to think about how such a forum can be produced and sustained. Most importantly, we would like to think that we have affected, in those small and significant ways, the intellectual community that sustains us.

We’ll let the students decide, wittingly or not, they always do.

There shall be a third issue next spring.

Adam Sassoon
Dave Shukla
# Contents

Some Requirements for a Successful Account of Sentential Content Identity  
*Tarak Shah*  
1

Supervaluationism and the Sider/Braun Theory of Vagueness  
*Kevin Salinger*  
19

Riddles of Induction  
*Aaron Butler*  
33

Two Externalist Theories of Perception?  
*Paolo Santorio*  
51

The Philosopher-King and the King-Philosopher  
*Aviv Rosenblatt*  
61

Room to Talk? 'Archeology and Experience' in Early Foucault  
*Erin Beeghly*  
85

The Perceptual 'Mirror' in Merleau-Ponty's Phenomenology  
*Mitchell Anderson*  
93

The Philosophical Question of Race and its Import in the Affirmative Action Debate  
*Julie Anne Genter*  
105

Ontological Privilege: In Defense of Natural Rights  
*Alexandra Khrebtukova*  
121

Nozick, Cohen, and Scheffler: Entitlement, Ownership, and a Meaningful Life  
*Noam Jacob*  
131

With Responses By:  
*Adam Sassoon*  
151

*David Jonas Cohn*  
157

*Dave Shukla*  
167

Notes on Contributors  
181
TARAK SHAH

Some Requirements for a Successful Account of Sentential Content Identity

"Clearly, there's no death. Saying what's clear to all is banal.
I'd like it explained to me.

~ Anna Akhmatova, "The Year 1913"
translated: D.M. Thomas
In §64 of *The Foundations of Arithmetic*, Frege introduces what is now often called the Direction Equivalence (D=):

\[ \text{the Direction of } a = \text{the Direction of } b \text{ iff } a//b \]

as a means of implicitly defining the concept of direction. This is actually only an instance of the direction equivalence, since the actual equivalence is a universally quantified sentence. Since much of my discussion is about content identity between sentences, I will restrict my discussion to instances, assuming that if content identity holds in every instance, then the universally quantified equivalence is acceptable. Frege argues that although we have an intuition of straight lines, we have no such intuition of the direction of a line, and so in order to discover the concept of direction we must begin with the concept (which we already have) of straight lines, and engage in some “intellectual activity.” So we are to take D= as giving us the concept of direction through our prior knowledge of parallel lines. Frege explains how this is achieved:

Thus we replace the symbol // by the more generic symbol =, through removing what is specific in the content of the former and dividing it between \(a\) and \(b\). We carve up the content in a way different from the original way, and this yields us a new concept.

Frege does not make more rigorous the notion of carving up content. Having a better understanding of it, though, would be useful. D= is supposed to be the analogue for directions of one proposal, now called Hume’s Principle (HP), for arriving at the concept of number:

\[ \text{The Number of F’s } = \text{The number of G’s iff there is a 1-1 correspondence between the F’s and the G’s} \]

As with D=, the two sides of this equivalence are, if the equivalence is to introduce numerical singular terms, to be taken as recarvings of the same content. So it would be useful to have a detailed account of “content,” and especially a notion of when two sentences can be said to express the same content. If, on such an account, we can show that the two sides of D= and of HP express the same content, we will have also taken an important step towards justifying Frege’s Platonism and towards the goals of his logicist program. In this paper I want to explore some of the requirements on such an account, and give a detailed discussion of how slingshot arguments af-
fect this project. Much of the paper is motivated by Bob Hale’s attempt to formulate an account of content identity in “Grundlagen §64” and the questions it gives rise to.

**Sentential and Relational Content**

Bob Hale, in his essay, “Grundlagen §64,” attempts to work out an account of when two sentences count as having the same content. After a lengthy postscript, he comes to what he considers a suitable relation of content-identity, the ancestral \( \phi^* \) of the relation \( \phi \) defined as follows:

\[
\phi(S_1, S_2) \text{ iff } S_1 = S_2 \text{ or } S_1 \text{ and } S_2 \text{ are compactly equivalent [compact entailment and compact equivalence will be explained below] or for some sentences } A, B \text{ and expressions } E, E' : A \text{ compactly entails } B \text{ and } S_1 \text{ and } S_2 \text{ result from } A \text{ and } B \text{ respectively by uniform bilateral [non vacuous] replacement of } E \text{ by } E' \quad (2001, \text{ pp. 113-115}).
\]

I will discuss this proposal in more detail below. However, it will be helpful to follow some of Hale’s reasoning before this point in order to get a better grasp on the problem and hopefully to shed light on possible solutions.

Hale begins by rightly pointing out that what Frege considers as being recarved in §64 is not the content of a sentence, but that of the symbol for parallelism. However, he still holds that an account of content-identity at the sentential level will be sufficient for the present purposes:

But it will bear indirectly, if we suppose, plausibly, that the content of a subsentential expression is part of the content of any complete sentence containing it, so that any recarving of the content of the former is *eo ipso* a recarving of the latter (2001, p. 93).

What he says here seems to be missing an important point. He has said that any recarving of the content of a subsentential expression (for our case, a relation symbol) is a recarving of the complete sentence containing it. This may be right, but it does not answer our question. What we want to do (and what Hale tries to do) is to define a relation of content-identity for sentences in such a way that it accounts for what it means to recarve the content of a relation symbol. Hale seems to have ignored the worry about how an ac-
count of content-identity for sentences will be able to explain recarvings of relational content. This worry is expressed by Potter and Smiley:

We agree with Hale . . . that success in recarving the content of a relation – which, as he observes, is what Frege actually had in mind – is likely to result in recarvings at the sentential level. The doubt is the other way round: how can a discussion purely at the sentential level deal with the global question behind the success or failure of the relational project? (2001, p. 336)

The problem is highlighted by the fact that certain equivalence relations are recarvable into identity relations between objects, but not all. So, in $\text{D}^r$, the relation of parallelism is recarvable into an identity relation, or in HP, the relation of one-one correspondence is recarvable into an identity relation, but in Frege’s Basic Law V from the Basic Laws of Arithmetic, the relation of extensional equivalence is not recarvable into an identity relation (since this Law leads to Russell’s paradox, it must not be true).

So, any account of content-identity for sentences that aspires to shed some light on the notion of recarving content in §64 of the Grundlagen must be able to answer this challenge. A reasonable test would be to see if it can recognize HP as an instance of recarving content, and also see that Basic Law V is not. An account could be successful without passing this test, but it would still need to explain the relationship between recarving the content of sentences and recarving the content of relation symbols.

Slingshots

Hale considers taking the content of a sentence to be the state of affairs it represents, but quickly discards the idea because of the difficulties posed by ‘slingshot’ arguments. It would be helpful to examine in more detail what exactly these difficulties are.

Slingshot arguments have been used by several philosophers (including Davidson) to argue against the idea that there are facts or states of affairs to which true sentences refer or which true sentences represent. The argument goes from assumptions which are presumably innocuous in the eye of someone who accepts an ontology of facts or states of affairs, to the conclusion that every true sentence refers to the same state of affairs (or fact). The assumptions for this argument, which I will call Slingshot$_1$, are (this version of the argument is adapted from Davidson (1967, p.118)):
(∀x)(x is the state of affairs represented by S ↔ x is the state of affairs represented by T) if

(i) T is obtained from S by substituting, in one or more places, a co-referring singular term (that is, replacing ‘α’ for ‘β’ when we have the truth of ‘α=β’),

or

(ii) S and T are logically equivalent

Now, for some sentence Q, since Q is logically equivalent to ‘(ιy)(y=0 ∧ Q) = (ιy)(y=0)’, we get that

(A) (∀x)(x is the state of affairs represented by Q ↔ x is the state of affairs represented by ((ιy)(y=0 ∧ Q) = (ιy)(y=0)).

Now, for any sentence R that is materially equivalent to (shares the same truth value with) Q, (ιy)(y=0 ∧ Q) and (ιy)(y=0 ∧ R) refer to the same object. So, substituting co-referring singular terms in (A), we get

(B) (∀x)(x is the state of affairs represented by Q ↔ x is the state of affairs represented by ((ιy)(y=0 ∧ R) = (ιy)(y=0)).

Now however, we know that ((ιy)(y=0 ∧ R) = (ιy)(y=0)) is logically equivalent to R, so we have

(C) (∀x)(x is the state of affairs represented by Q ↔ x is the state of affairs represented by R).

Since Q and R were arbitrary materially equivalent sentences, we have shown that all true sentences represent the same state of affairs, as do all false sentences.

We might question this argument for several reasons. For one thing, the move from (A) to (B) seems to require definite descriptions to be taken as referring singular terms (as the move invokes the substitutability of co-referring singular terms in order to replace one definite description with another that purports to describe the same individual). At least, it is conceivable that, for the inference from (A) to (B) to work, we might need to add to (i) that
not only can co-referring singular terms be substituted, but so can definite
descriptions that describe the same individual. Furthermore, we may find
(ii) suspicious. Why should all logically equivalent sentences represent the
same state of affairs? This seems to be what Hale wants to avoid, when he
claims (2001, p. 99) that arbitrary necessary truths “are not plausibly taken
as reconceptualizations of [the] same state of affairs.”

Rejecting (ii) will not help us in the end, since, as Neale has pointed
out (1995, pp. 776-780), there is a version of the slingshot which does not
require (ii). Neale attributes the argument to Gödel, and the assumptions
required for this version of the argument are (Neale presents it using facts,
but the same argument can be made for states of affairs):

(I) ‘φ(a)’ and ‘a = (ιx)(x=a ∧ φ(x))’ stand for the same fact.

(II) Any sentence that stands for a fact can be put into
predicate-argument form (so, for example, “Socrates
snored and Plato snored” can be read as the predicate “x
snored and Plato snored” with the argument “Socrates”).
Without this assumption, the argument can still be used to
show that all true atomic sentences refer to the same fact.

(III) In Gödel’s words, “the signification of a com-
plex expression, containing constituents which
have themselves a signification, depends only on
the signification of these constituents (not on the
manner in which this signification is expressed).”

(IV) ‘F((ιx)(φ))’ and ‘Fa’, where ‘a’ refers to the
unique object satisfying φ, stand for the same fact.

The argument goes:

Assume that (1), (2), and (3) are true:
(1)  Fa
(2)  a≠b
(3)  Gb

Then, by assumption (I), (1) stands for the same fact as

(4)  a = (ιx)(x=a ∧ Fx).

Similarly, (2) stands for the same fact as
(5)   \( a = (\exists x)(x=a \land x \neq b) \).

Now, because of assumption (IV), (4) and (5) stand for the same fact (the fact that \( a = a \)). Next, again by (I), we see that (3) stands for the same fact as

(6)   \( b = (\exists x)(x=b \land Gx) \).

Likewise, (2) stands for the same fact as

(7)   \( b = (\exists x)(x=b \land x \neq a) \).

As before, because of (IV), (6) and (7) stand for the same fact. But what we have shown is that (1) and (4) stand for the same fact, (2) and (5) stand for the same fact, and (4) and (5) stand for the same fact. By transitivity, then, (1) and (2) stand for the same fact. We have also shown that (3) and (6) stand for the same fact, (2) and (7) stand for the same fact, and (6) and (7) stand for the same fact. Again by transitivity, (2) and (3) stand for the same fact. In all, then, (1), (2) and (3) stand for the same fact, and in particular, (1) and (3) stand for the same fact.

Now, if instead of (2) above we had

(2*)   \( a=b \)

we could still run a similar argument. (1) and (4) would still stand for the same fact, while (2*) would stand for the same fact as

(5*)   \( a = (\exists x)(x=a \land x=b) \)

(2*) would also stand for the same fact as

(7*)   \( b = (\exists x)(x=b \land x=a) \).

As before, (3) and (6) still stand for the same fact. Again, because of (IV), (4) and (5*) stand for the same fact, as do (6) and (7*). So (1) and (2*) stand for the same fact, and (2*) and (3) stand for the same fact, so again (1) and (3) stand for the same fact. So we have shown that every true sentence stands for the same fact.

Notice that this version of the slingshot uses an assumption about substituting singular terms for definite descriptions, and vice versa, unlike in the case of Slingshot1, where we treated definite descriptions as singular terms.
and just used an assumption about substitution of co-referring singular terms. So far in this presentation of Gödel’s version we have remained ambivalent about whether definite descriptions are singular terms, but if they are, then the rule (IV) would be equivalent to (i). What happens if instead we treat definite descriptions as Russell did, that is, we take $\phi((\iota x)\psi(x))$ to be short for

$$(\exists x)((\forall y)(\psi(y) \leftrightarrow y=x) \land \phi(x))$$

Now we seem to be in a position to deny the step from (A) to (B) in Sling-shot$_1$. We do not have a sentence in the form ‘$\alpha=\beta$’ that would allow us to carry out our substitution; the statement we are relying on:

$$(\iota y)(y=0 \land Q) = (\iota y)(y=0 \land R)$$

does not actually have the form of an identity, but of an existential assertion:

$$(\exists x)((\forall y)((y=0 \land Q) \leftrightarrow y=x) \land (\exists w)((z=0 \land R) \leftrightarrow z=w) \land w=x))$$

So the step is not justified by (i). In fact, as Neale points out, if we adopt a Russellian interpretation of definite descriptions, we see that Slingshot$_1$ depends not on assumptions (i) and (ii) but actually on (ii) and (i*) ‘T’ is obtained from ‘S’ by substituting, in one or more places, a definite description for a definite description when both are satisfied by the same unique object.

As far as Gödel’s slingshot goes, if we adopt Russell’s treatment of definite descriptions, then we see clearly that (IV) is not equivalent to (i) and so we can reject (IV) without rejecting (i).

This is all to say that Hale’s rejection of content as state of affairs may have been premature. After all, we were, from the beginning, in a position to reject Slingshot$_1$ because of our reservations about (ii). In addition, adopting a Russellian treatment of definite descriptions, we can also question the admissibility of (i*). That is, we might argue that ‘$F((\iota x)(\phi))$’ and ‘$F((\iota x)(\psi))$’ when $\phi$ and $\psi$ are satisfied by the same unique object, do not in fact represent the same state of affairs. For example, do
(8) ‘The Thane of Cawdor killed Duncan’ and

(9) ‘The murderer of Duncan killed Duncan’

represent the same state of affairs? Intuitively, it seems that (9) could represent a state of affairs in which (8) is false – the case where The Thane of Cawdor ≠ The murderer of Duncan.

So we might want to say (8) and (9) do not represent the same state of affairs, and so we can reject (i*), and in this way hold that the content of a sentence is the particular state of affairs it represents without being vulnerable to Slingshot.

Similarly, we can avoid Gödel’s slingshot by rejecting (IV). After all, for the same reasons as above, we ought to deny the claim that (8) and

(10) ‘Macbeth killed Duncan’

represent the same state of affairs. In developing a notion of when two sentences represent the same state of affairs, then, it seems we ought only to allow substitutions involving “rigid designators.” Kripke (1972, p. 258) defines a rigid designator as a designator that designates the same object in every possible world. While he argues that all names are rigid designators, there are in addition descriptions that designate rigidly. Consider, for example, ‘The successor of three’. Though this is a definite description, it designates rigidly: the successor of three is four in every possible world.

So we know that we need to reject (i*) and (IV) in order to avoid slingshot arguments, but we cannot easily deny that ‘The successor of three is a multiple of two’ and ‘The predecessor of five is a multiple of two’ represent the same state of affairs, as we could with (8) and (9).

Perhaps, then, in the case of rigidly designating descriptions, we should allow substitution – treat these descriptions as singular terms – but not allow it for any other definite descriptions. Looking back at Gödel’s slingshot, we might amend (IV) according to the suggestion, so that it reads

(IV*) ‘F((ιx)(φ))’ and ‘Fa’, where a is the unique object satisfying φ, stand for the same fact whenever ‘(ιx)(φ)’ rigidly designates the object a.

Indeed, this would block Gödel’s slingshot, since the descriptions that are involved in the substitutions do not designate rigidly (e.g. ‘(ιx)(x = a ∧ Fx)’). Furthermore, it is not clear to me that there is a version of the slingshot that could go through with only the weaker rule (IV*) instead of (IV). So, ten-
tatively, although we are rejecting (IV), we can still accept (IV*). Hale does seem to accept at least this much substitution. For example, when giving an example to prove that φ is not transitive (which is why he eventually settles on the ancestral of φ), he claims that ‘6 is greater than the number of moons of Mars’ and ‘The successor of 5 is greater than the number of moons on Mars’ express the same content (2001, p. 114).

A theory of content identity that restricts substitution according to whether or not a description is a rigid designator, while interesting, would require a longer discussion of rigid designators, including some clarification of the notion of possible worlds. As things stand, we will see that Hale will not be able to take this route anyways, since his formulation is vulnerable to substitutions beyond (IV*).

So, if we are somehow able to take the content of a sentence to be the state of affairs it represents, how should we determine when two sentences represent the same state of affairs? We have ruled out some suggestions, and it also seems reasonable to accept certain suggestions, such as (i), (I), and (III) above. But these must be minimal conditions on whatever account we are to give; they tell us what sorts of substitutions are admissible, but they do not explain why the two sides of D= represent the same state of affairs, and so the same content.

Before reviewing Hale’s proposal, I want to quickly summarize a couple of key points that we have made about any successful account of sentential content identity.

(R1) Any such account, if it is to be at all useful for the purposes of Frege’s project – that is, for the purposes of justifying implicit definitions, specifically HP, has to explain how recarving at the sentence level is related to recarving at the relational level. An account exact enough to find HP to be an instance of content identity while also finding Basic Law V not to be one could help to explain this relationship.

(R2) If we intend to use identity in states of affairs represented as what content identity of sentences consists in, we have to reject rules such as (i*) and (IV) above. More generally, in Neale’s terminology, we are going to have to reject ι-substitution rules (these include substituting one description for another, substituting a singular term for a description, and substituting a description for a singular term) within the
I should make one last point before moving on. By rejecting the notion that the content of a sentence is the state of affairs it represents, Hale is not immune to the slingshot arguments, and so cannot ignore requirement (R2). This is because the slingshot arguments do not only attack the notion of states of affairs. If we replace the connective ‘the state of affairs represented by $x = \text{the state of affairs represented by } y$’ with ‘the content of the sentence $x = \text{the content of the sentence } y$,’ we have not freed ourselves of any burden.

**Compact Entailments**

Because of the possible vulnerability to slingshot arguments, Hale does not at first try to define the content of a sentence as the state of affairs it represents. Instead, he defines sameness in truth condition, and uses a “derivative” notion of states of affairs – two sentences represent the same state of affairs if and only if they have the same truth condition (2001, p. 102) (notice again, however, that ‘$x$ has the same truth condition as $y$’ also must answer to requirement R2, since the slingshot argument is still applicable). So recall that Hale formulates the relation (of sameness in content, or sameness in truth condition, or sameness in pseudo-states of affairs) as the ancestral $\phi^*$ of the relation $\phi$ defined between $S_1$ and $S_2$, where $S_1$ bears $\phi$ to $S_2$ whenever:

1. $S_1 = S_2$
2. $S_1 \iff S_2$
3. For some sentences $A, B$ and expressions $E, E'$: $A \Rightarrow B$ and $S_1$ and $S_2$ result from $A$ and $B$ respectively by uniform bilateral replacement of $E$ by $E'$ (2001, pp. 113-115).

$\Rightarrow$ and $\iff$ stand, respectively, for *compact entailment* and *compact equivalence*. Compact equivalence is defined as mutual compact entailment, and compact entailment is defined as follows (2001, p. 112):

$A \Rightarrow B$ iff

- A entails B and
- For every non-logical expression $E$ in $A$, there is some $E'$ such that $A(E'/E)$, i.e., the result of substituting
E′ for E uniformly throughout A, does not entail B and (c) For every subformula S of A, there is some modal equivalent S′ such that A(S′/S) does not entail B.

(He takes ‘entails’ to include not only logical entailments, but also “analytic” or “conceptual” entailment – that is, the case when ‘A ∧ ¬B’ is logically possible, but not conceptually possible, for example, ‘this colored patch is yellow’ entails ‘this colored patch is not red’). By saying two sentences are “modally equivalent,” Hale means that they have the same modal status – they both express impossibilities, or both necessities, or both contingencies.

Now, (φ3) in the formulation of content-identity seems problematic, since it introduces directionality into what ought to be a symmetrical relationship. Indeed, we will soon see what sorts of problems this can lead us into. So instead I will use the modified version of (φ3):

(φ3∗) for some sentences A, B and expressions E, E′: A⇔B and S1 and S2 result from A and B respectively by uniform bilateral replacement of E by E′.

This modification does not create new problems, but does resolve some difficulties in the formulation.

Hale says he is able to avoid Slingshot1 by rejecting both (i) and (ii) (2001, p.103). At first glance, it is difficult to see how he has rejected (i). After all, sentences such as ‘Rba ∧ a=c’ and ‘Rbc ∧ a=c’ are compactly equivalent and so, according to Hale, express the same content. It is only when we include definite descriptions as singular terms that we can find a case where (i) does not hold. Specifically, we can use the example of step (B) in Slingshot1. We want to show that Q and (((ιy)(y=0 ∧ R)=(ιy)(y=0)) – where Q and R are materially equivalent – do not express the same state of affairs. That they fail (φ1) is straightforward. Since there is no uniform replacement of ‘0’ in (((ιy)(y=0 ∧ R)=(ιy)(y=0))) that disrupts the entailment from that sentence to Q, that sentence does not compactly entail Q, so even though Q compactly entails (((ιy)(y=0 ∧ R)=(ιy)(y=0))), the two are not compactly equivalent. Finally, the two sentences have no semantically significant parts in common, so (φ3∗) is not a possibility. So we have shown that, as long as we take definite descriptions as singular terms, Hale’s formulation rules out (i).

To see how Hale rejects (ii), notice that he can claim that ‘((ιy)(y=0 ∧ Q) = (ιy)(y=0))’ does not express the same content as ‘Q.’ (φ1) fails obviously,
(φ2) fails because there is no uniform replacement for ‘0’ in ‘(ty)(y=0 ∧ Q) = (ty)(y=0)’ that results in a sentence that does not entail Q, and (φ3*) fails because there are no such A and B. Notice, however, that, depending on how we phrase the question, (φ3) may not have failed. If we ask if ‘Q’ expresses the same content as ‘(ty)(y=0 ∧ Q) = (ty)(y=0),’ then by (φ3) we would have had to say yes. Let A be R, and let B be ‘(ty)(y=0 ∧ R) = (ty)(y=0),’ and now the two sentences are obtained from A and B by uniform bilateral replacement of ‘R’ by ‘Q’. Since A ⇒ B, (φ3) is satisfied.

In any case, Hale claims that he can reject both (i) and (ii). As we have seen, though, the claim that he is able to avoid Slingshot1 by rejecting (i) suggests that Hale is reading definite descriptions as singular terms. What if he instead takes a Russellian position? Is Hale then in a position to reject (i*)? That is, can he reject that ‘(∃x)((∀y)(ψ(y) ↔ y=x) ∧ φ(x))’ and ‘(∃x)((∀y)(ω(y) ↔ y=x) ∧ φ(x))’ express the same content whenever ω and ψ are satisfied by the same unique object? This last assumption can be expressed as ‘((∃x)((∀y)(ψ(y) ↔ y=x) ∧ (∃w)((∀y)(ω(y) ↔ y=w)) ∧ w=x))).’ So in all, we want to know if

\[1\] (((∃x)((∀y)(ψ(y) ↔ y=x) ∧ (∃w)((∀y)(ω(y) ↔ y=w)) ∧ w=x))) ∧ (∃x)((∀y)(ψ(y) ↔ y=x) ∧ φ(x))

and

\[2\] (((∃x)((∀y)(ψ(y) ↔ y=x) ∧ (∃w)((∀z)(ω(y) ↔ z=w)) ↔ w=x))) ∧ (∃x)((∀y)(ω(y) ↔ y=x) ∧ φ(x))

express the same content. First of all, [1] and [2] do entail each other. In the two sentences, ψ, φ, and ω are the only non-logical constituents. Each one of them can be replaced uniformly in [1] to ruin the entailment from [1] to [2]. Similarly, we can always ruin the entailment from [2] to [1]. Every subformula in both [1] and [2] is contingent, so it is not difficult to come up with modally equivalent replacements which ruin the entailments in both directions. Therefore, [1] and [2] are compactly equivalent, and so by Hale’s criteria they express the same content. So he can not deny (i*). To avoid Slingshot1, then, Hale is forced into reading definite descriptions as singular terms.

Now, taking definite descriptions as singular terms, we see that Hale cannot reject (IV). It is pretty straightforward to show that (4) and (5) in the above presentation of Gödel’s slingshot, as well as (6) and (7), without performing any Russellian analysis, are compactly equivalent, and so should represent the same content. Similarly, Hale is unable to reject (I), and so as long as he takes definite descriptions as singular terms, Hale’s formulation is vulnerable to Gödel’s slingshot.
The situation now is this: either Hale accepts a Russellian reading of definite descriptions, in which case Slingshot becomes a problem, or he reads definite descriptions as singular terms, in which case Gödel’s slingshot creates problems. In either case, Hale’s formulation of content identity is reduced to one such that every true sentence expresses the same content.

So we have shown that Hale has failed to satisfy (R2). Whether or not he has satisfied (R1) is a more difficult question. Let’s try first to apply Hale’s relation in the case of $D^=\,\epsilon$. Our question is whether $D^=\,\phi^*$ to $D^=\,\phi^*$:

\begin{align*}
&D^=\,\phi^* \quad \text{D= (left): The Direction of a = The direction of b} \\
&D^=\,\phi^* \quad \text{D= (right): a//b}
\end{align*}

$(\phi_1)$ obviously does not do the trick. $(\phi_2)$ seems a bit problematic. The first question we might ask when trying to decide whether or not $D^=\,\phi^*$ is true is whether it is true that $D^=\,\phi^*$ entails each other (the first conjunct in the definition of compact entailment). A positive answer to this presupposes the truth of $D^=\,\phi^*$. In any case, let’s assume that the two sides do, in fact, entail each other. Taking the definite descriptions ‘the direction of a’ and ‘the direction of b’ as singular terms, we can look at $D^=\,\phi^*$ and notice that we can uniformly substitute ‘c’ for ‘a,’ as we can ‘c’ for ‘b,’ and ‘>’ for ‘//’ and in each case the resulting sentence does not entail $D^=\,\phi^*$. Indeed, even if we take the descriptions as Russellian (as long as we still assume that the two sides entail each other), we can still perform the same substitutions on $D^=\,\phi^*$ and achieve the same results. Since $D^=\,\phi^*$ has no proper subformulae, conjunct (c) in the definition of compact entailment need not worry us. So we have shown that, by $(\phi_2)$, $D^=\,\phi^*$ expresses the same content. We can do the same for HP. The only questionable point is whether we are justified in assuming that either ‘$d^=d^\prime$’ (that is, when the descriptions are taken as singular terms), or ‘$(\exists x)(\exists z)((\forall y)(y=Direction(a) \leftrightarrow y=x) \land (\forall w)(w=Direction(b) \leftrightarrow w=z) \land x=z)$’ entail and are entailed by ‘$a//b$’. It seems we would need a better notion of “conceptual” entailment. After all, if we were already in a position to say that ‘$D^=\,\phi^*$’ or ‘$D^=\,\phi^*$’ are conceptual impossibilities (they are not impossibilities in classical logic), then this entire discussion is pointless – explaining sameness of content was supposed to shed light on how $D^=\,\phi^*$ or HP is an instance of content recarving.

Now, our question is whether or not Hale’s definition rules out Basic Law V from counting as expressing the same content. An instance of Basic Law V:
As far as compact equivalence goes, the discussion above applies equally to this case. That is, BLV passes the uniform substitutability rule, since we can replace for both ‘F’ and for ‘G’ on the right side, and in both cases end up with a sentence that does not entail the left side. So the only question is whether or not the two sides mutually entail each other. Again, we need a better notion of “conceptual” entailment. If there is a way of explaining conceptual entailment that also explains why HP is an instance while BLV is not, then the ability to explain recarving the content of a relation symbol solely in terms of recarving the content of a sentence would be a little less doubtful. Otherwise, (R1) is still a concern.

2 is odd or London is a city – “conceptual” entailment or impossibility

In the postscript to his essay (2001, p.111), Hale, when considering some of the problem entailments that he would like to rule out in his account, considers

(*) 2 is odd \lor \text{London is a city} \text{ entails} \text{ London is a city.}

Since ‘2 is odd’ is, in Hale’s words, only a “passenger,” he does not want to count (*) as a compact entailment. He suggests that “it should no more qualify as compact than should instances of the schema: \( (p \land \neg p) \lor q \text{ entails q} \)” (111). This suggestion may give us some sort of hint as to how to understand “conceptual” entailment (I should mention that Hale himself only talks about “conceptual impossibility,” but the question is still the same: what are our criteria for saying A entails B when the entailment is not logical?).

The suggestion that Hale makes draws a parallel between \( (p \land \neg p) \) and ‘2 is odd.’ Clearly, we are to take ‘2 is odd’ to be an impossibility, but it is not a logical impossibility. So the question is, how do we treat conceptual impossibilities, when we are working in systems that only give us definite answers about logical impossibilities? The parallel Hale draws here, then, suggests that we should treat conceptual impossibilities in exactly the same way as we treat logical ones. Even if this is the correct treatment, it still requires that we already know whether or not something is a conceptual impossibility. In the case of HP or D=, this requirement leaves too much of the work to the still very vague topic of conceptual impossibility, which Hale does not really explain.
Recarving

Through an examination of Hale’s arguments and especially of Sling-shot arguments and definite descriptions, we have highlighted some of the difficulties overlooked by Hale in understanding §64 of the *Foundations of Arithmetic*. (R1) and (R2), for example, must be addressed by anyone wishing to give a satisfactory account of content recarving, and in order to do this a notion of conceptual impossibility needs to be worked out, including a more thorough discussion of how logically to treat conceptual impossibilities, and some sort of criterion for what counts as a conceptual impossibility. Furthermore, the above discussion of rigid designators may open up the possibility of a formulation of content-identity that is not vulnerable to the various slingshot arguments.

Even if these issues are worked out, it is not clear that we would have justified Platonism. For one thing, as Frege himself points out, the Caesar problem would still remain (this is in fact why he abandons the notion of carving up content in the first place). Furthermore, Potter and Smiley (2001, pp. 337-8) have suggested that the best account we can give of carving up content still falls short of giving us the existence of objects such as directions or numbers, and so anyone giving such an account would have to explain how the existence of these objects is secured. Clearly, then, there is much more to be done than to give an account of content recarving. But such an account would be an important step, and this discussion has outlined some constraints and some challenges to which such an account would have to respond.

WORKS CITED


“When I say 'the book is lying on the table,' does this really have a completely clear sense? This is a very important question.”

~ L. Wittgenstein, *Notebooks, June 20, 1915*
I. Introduction

The standard sorites paradox is well known:

(P1) 1 grain of sand is not a heap.

(P2) For each \( n \), if \( n \) grains of sand are not a heap, then \( n+1 \) grains of sand are not a heap.

From P1 and P2, by repeated application of modus ponens (or, if you like, by a single use of both mathematical induction and universal instantiation), we infer:

(C) 1,000,000 grains of sand is not a heap.

Given the plausibility of P1 and P2, and the absurdity of C, it seems that we have deduced a false conclusion from true premises. Thus, one of the premises must be false or the mode of inference invalid. Yet P1, P2, and modus ponens are each highly plausible. The situation is not unique to ‘heap’ – virtually any vague predicate (‘tall’, ‘red’, ‘funny’, etc.) is susceptible to a paradox of this kind. Attempts to resolve the paradox typically are subsumed in a broader theory of vagueness – one that purports to justify rejecting P1, P2, or modus ponens.

In a recent paper\(^1\), Ted Sider and David Braun propose an account of vagueness, based in some respects on both the supervaluationist and Fregean approaches, which purports to resolve the sorites paradoxes while avoiding the problems of supervaluationism. Their leading idea is that no sentence containing a vague word is, strictly speaking, true (or false). Thus, both premises in a sorites paradox turn out untrue, and so the argument is unsound. They locate their theory within the typical supervaluationist structure, arguing that they can retain its positive aspects while also solving its problems. In this paper, I argue that their account does not live up to its motivations: the Sider/Braun theory fails to solve key problems facing supervaluationism, presents new problems of its own, and in general makes no significant progress over supervaluationism.

They begin by adopting something like Keefe’s supervaluationist semantics\(^2\). They understand the phenomenon of vagueness as a case of semantic deficiency or indeterminacy. For any vague sentence, we can, in principle, make the sentence precise in a number of distinct ways by providing a precise extension to each vague predicate and a precise reference to each vague name occurring in the sentence. In doing so, we form a set
of admissible precisifications of the sentence. Moreover, as Keefe writes, for any predicate $F$, “no single one of $F$’s precisifications can be correctly identified as providing the extension of $F$, because the meaning of vague $F$ is such as to leave the choice between them unsettled. Not only have we not consciously made the choice between these alternative extensions, but… nothing about the world, or about our use of the word, picks out a unique extension” (155). Sider and Braun accept this starting point.

On a supervaluationist semantics, a sentence is true simpliciter if and only if the sentence is supertrue – i.e., true on all legitimate precisifications (similarly for false sentences). If a sentence is true on some legitimate precisifications and false on others, then it is truth-valueless. Sider and Braun reject the equation of truth simpliciter with supertruth: they claim that the notions of truth and falsity only make sense relative to a precisification, and hence vague sentences should not receive a truth value at all. Truth and falsity are technical semantic notions that can only be applied to completely precise sentences. This claim has its roots in the position they attribute to Frege, that logic and semantics apply to only precise languages, and thus virtually never to natural languages. For example, the sentences, “Michael Jordan is tall,” “Black dogs are black,” “Mt. Everest is a mountain or Mt. Everest is not a mountain” and even, for some perfectly precise sentence $\phi$, “$\phi$ and not-$\phi$ and John is tall,” do not receive a truth value, in virtue of the fact that they contain vague names or predicates. Arguments like, “All men are mortal, Socrates is a man, therefore Socrates is mortal,” are unsound, because the premises are not true. Sider and Braun realize that this sounds like a radical position, but they have arguments designed to make the claim more palatable.

II. Why do Vague Sentences Lack a Truth Value?

A. Analogy to Physics

Sider and Braun draw an analogy to idealizations in physics. Scientific theories are often developed in idealized worlds that bracket noise such as friction, air resistance, dark matter, etc., so that the theory may be kept simple, while still yielding powerful predictive and explanatory results. A consequence of this method, however, is that the theory will, strictly speaking, only apply in the real word to situations that are free of the so-called bracketed noise. They think that they can tell a similar story for classical logic and semantics. Introductory logic books start out by specifying a number of idealizations on the object language: each name refers to a unique
object in the domain, the extensions of relation symbols are precise, and so on. It is unreasonable to expect that such theories would strictly apply to most natural language sentences, because the vast majority of them do not share the assumed idealizations. Therefore, they argue, classical semantics, and in particular the notions of truth and falsity, strictly speaking, apply only to perfectly precise sentences.

However, when the analogy is fully developed, it is not clear that the argument produces the desired results. Idealized scientific theory, while precluding the strict application of idealized theories to the real world, does not preclude the application of other physical concepts therein (mass, charge, etc.) to reality. If we accept Sider’s analogy, then it seems plausible that even if we see classical logic as an idealized theory, we might still be able to apply the notion of truth to everyday sentences. In other words, the analogy between idealization in science and logic might show that theories in each domain involve idealizations, but he has not yet shown that truth is one such idealization and therefore an inapplicable concept.

Even if they can show that truth, as used in classical semantics, involves such an idealization, there is a stronger objection that I do not believe the analogy can overcome. In science, it is true that theories are initially developed by idealizing away a significant level of noise that shows up in the real world, but the story has not been finished. A prominent goal of science, and likewise of logic and semantics (with respect to natural languages), is to give an account of the full range of available data. Toward that aim, simplified theories are refined, expanded, and developed to take into account the previously bracketed variables. When a more inclusive and explanatory theory is developed, it becomes the new standard, in place of the previous theory. This method holds for both science and logic. Consider the liar. When its implications were fleshed out, people did not chalk up self-referential sentences or truth predicates as a feature of natural language that must be idealized away before logic and semantics could be done. If they did, no progress would ever have been made. So, in the case of vagueness, it seems that the question is whether it seems reasonable to generalize beyond classical logic and semantics to logic and semantics in general, and claim that they simply do not apply to vague language. The analogy to physics does not answer that question. At best, it suggests that we should focus our attention, as many philosophers have done, on developing more powerful logics and semantics that can succeed where our current theories fail.

B. Analogy to Ambiguity

Their second argument for the claim examines the way that we handle
sentences containing ambiguous words. They argue that ambiguity, like vagueness, is a kind of semantic indeterminacy. They claim that ambiguous sentences, whose meaning is not made clear by context, are not true or false simpliciter, but only true or false relative to a disambiguation. They don’t give much of an argument for that point, but it seems right: suppose that we wanted to assign a truth value to the sentence, ‘Kent went to the bank’, where it is unclear which sense of ‘bank’ is being used. If we insist on assigning a truth value to the sentence, we might follow Fine’s strategy and interpret it using a supervaluationist technique. Consider:

(1) Kent went to the bank
(2) Kent went to the financial bank \land Kent went to the river bank
(3) \neg Kent went to the financial bank \land \neg Kent went to the river bank

We might say that (1) is true iff (2) is true, (1) is false iff (3) is true, and (1) is neither true nor false otherwise. However, if (1) was asserted by a speaker with a particular sense of ‘bank’ in mind, this interpretation is unable to guarantee the correct truth-value of the assertion. Suppose that a speaker utters (1), intending to mean that Kent went to the financial institution. Suppose that, in fact, Kent did go to the financial institution, and not the river slope. If we follow Fine, the speaker’s utterance would be neither true nor false, while we intuitively want to say that the utterance was true. Similar counterexamples arise for other truth-function ways of assigning truth-values to ambiguous sentences. Thus, it seems correct that ambiguous sentences should not receive a truth-value before disambiguation, at least when the speaker has a particular sense of the word in mind. Sider and Braun observe that the reason why we do not assign a truth-value to ambiguous sentences is that they are semantically indeterminate. But, it is central to their, and the supervaluationist, approach that vagueness is a type of semantic indeterminacy as well. He argues that, because vagueness and ambiguity are, in that way, closely related phenomena, vague sentences should be treated in the same way as ambiguous sentences: just as ambiguous sentences do not receive a truth value before disambiguation, vague sentences do not receive a truth-value before precisification.

However, upon scrutiny, a relevant disanalogy emerges. While vagueness may be a kind of semantic indeterminacy, ambiguity is apparently a case of semantic deficiency, not indeterminacy. On Sider and Braun’s view, there are a number of precise meanings associated with a vague sentence, but there is no fact of the matter (i.e., it is indeterminate) as to which one of those meanings the sentence picks out. Yet in the case of ambiguity, a
speaker almost always intends to use the ambiguous word in a determinate sense, merely failing to communicate that sense in the utterance\(^7\). The disanalogy is relevant because it is for the very reason that speakers use ambiguous words in determinate senses that I reject Fine’s supervaluationist approach to ambiguity. If ambiguity, like vagueness, were a case of semantic indeterminacy, then it is controversial whether Fine’s strategy would fail in the way Sider and Braun’s theory requires that it does.

C. Vague Sentences Fail to Express a Proposition

The third argument goes something like this: a sentence is true just in case it expresses a unique and true proposition. But, a vague sentence fails to pick out a proposition; it merely points to a cloud of propositions. Each proposition in the cloud is perfectly precise and bears a truth value, but it is indeterminate which proposition is expressed by the vague sentence. Sentences which do not pick out a proposition do not, in general, receive a truth-value (e.g., ‘Twas brillig, and the slithy toves did gyre and gimble in the wabe’), and so neither do vague sentences.

The key premise of the argument is that propositions are necessarily precise. Yet, it is not obvious that they must be precise. They have no supporting arguments for the claim that propositions necessarily are. Nor is it unquestionably assumed in the philosophy of language that propositions cannot be vague. It does not seem that Sider can rely on this argument without opening a Pandora’s Box of issues philosophical issues surrounding propositions – a debate that they choose to avoid. This speaks strongly in favor of classical supervaluationism, which, as Keefe points out, “can be conducted entirely without reference to propositions” (158). They might defend the argument by saying that ‘proposition’ is a technical term, and that they can simply define them to be precise. However, on that approach, the premise that ‘a sentence is true just in case the proposition it expresses is true’ starts to look dangerously question-begging.

III. Solution to the Sorites

If Sider and Braun are right, then we have a natural solution to the sorites: the premises of the argument contain occurrences of vague words, so they do not receive a truth value. Thus, the sorites argument is unsound, as its premises are not true. To make them true, we would have to make them precise. But, in doing so, we would create a sharp boundary, and so the second premise, \(\forall n(F(n) \rightarrow F(n+1))\), is false – the argument is still
unsound. Paradox resolved.

IV. Assertability Conditions and the Semantics of $\exists$, $\lor$

So far, Sider and Braun have argued that a Fregean-like approach to vagueness is palatable and successful at resolving the sorites. At this point, there is an objection that can be raised against their, and most Fregean, accounts of vagueness. Consider the intuitive principle that you should be willing to assert a sentence $\varphi$ only if you believe that $\varphi$ is true, i.e., truth is a norm of assertion and affirmation. Sider and Braun are already committed to the position that, if $\varphi$ is vague, then $\varphi$ is not true. Hence, if $\varphi$ is vague, we should not believe that $\varphi$ is true, and therefore we should not assert or assent to $\varphi$. So, we should not make vague assertions. Yet, nearly all natural-language sentences are vague (including virtually every sentence in Sider and Braun’s paper). Unless they can say why we should assent to vague sentences, their theory will reduce us to near silence – undermining its own assertability in the process.

Truth is not a norm of assertability in the proposed theory. To understand exactly what they think the norm of assertability is, we need to examine their notion of approximate truth. They believe that they have already shown that vague sentences lack truth-values, whereas the precisifications of vague sentences do not. But, they also believe that our everyday utterances are true, vagueness ignored, or, a harmless approximation of the truth. To give the semantics for ‘approximate truth’, they need to locate it somewhere in their semantic framework (i.e., the supervaluationist framework). They take the obvious choice: a sentence $\varphi$ is approximately true just in case, on every admissible precisification, $\varphi$ is true, i.e., approximate truth is supertruth. Approximate falsity is superfalsity.

With this notion in hand, they may have a plausible candidate to substitute for truth as the norm of assertability. As a first shot, if we adopt approximate truth as the criterion, we end up with some nice results:

- Clear cases pass the test for assertability. For example, ‘Mt. Everest is larger than an ant,’ is true on every admissible precisification, hence it is approximately true, hence assertable.
- Borderline cases fail the test for assertability. For example, ‘A 5’10” tall man is tall,’ is true on some admissible precisifications and false on others, hence it is not approximately true, hence we should refrain from asserting it.
Sentences involving penumbral connections are generally treated in line with our intuitions. For example, ‘If 50 grains of sand are a heap, then 51 grains of sand are a heap,’ is approximately true, hence assertable.

So far, this looks a lot like supervaluationism. So, consider a sentence that supervaluationism notoriously has trouble with:

(4) There exists an \( n \), such that \( n \) grains of sand do not make a heap and \( n+1 \) grains of sand do make a heap.

On their view, (4) is approximately true, hence assertable. Yet, they cannot give an approximately true answer to the question, ‘Which \( n \)?’ for there is no consistent cut-off point across the set of admissible precisifications. A similar objection can be raised involving a disjunction in the place of an existential quantifier. Approximate truth, like supertruth, is elusive. Sider and Braun agree that this is a significant problem: if approximate truth is the norm of assertion, then we should be willing to assert (4). But, because no number satisfies (4) across every admissible precisification, we cannot assent to any particular instance of (4), and given the meaning of the existential quantifier, that seems to amount to a denial of (4). Thus, approximate truth cannot be the (sole) norm of assertion.

So, what must be added? Their answer seems to be something like this: for any vague sentence \( \varphi \), we should be willing to assert that \( \varphi \) just in case both \( \varphi \) is approximately true and the vagueness of \( \varphi \) is being ignored. The vagueness of sentences like (4) is “manifest” in virtue of our inability to pick out an approximately true instance of the sentence. Thus, they claim, we cannot both assert (4) and ignore the vagueness of ‘heap’, thus (4) fails the test for assertability.

It’s not clear, however, that their criterion for assertion is tenable. It seems as though we are perfectly willing to assert clear cases of vague sentences, even if we are not ignoring the vagueness involved (whatever that means – they fail to give an satisfying explanation of what it would be to ‘ignore the vagueness’ of a sentence). For example, if Timothy Williamson is teaching a course on the subject and explicitly not ignoring the vagueness of sentences like ‘Jordan is tall,’ ‘black dogs are black,’ ‘a man with no hair is bald,’ then, under Sider and Braun’s criterion, it seems that Williamson should refrain from asserting those sentences. What a peculiar thing to say! No matter how much I focus on the vagueness of the words ‘black’ and ‘dog’, I doubt that I will ever find myself unwilling to assert,
‘black dogs are black’, even if I accept that it does not count as being strictly true, but merely approximately true.

It will be helpful in evaluating Sider’s overall theory to contrast his account of (4) with Keefe’s. Sider takes it that certain sentences involving existential quantifiers or disjunctions pose a threat to supervaluationism. Yet Keefe has a reply to this problem, namely that the supervaluationist interpretations of ∃ and ∨ differ from classical logic only with respect to borderline cases and, as such, may actually be consistent with our intuitions. She admits that her response might be slightly counterintuitive, but she thinks that any theory of vagueness needs to accept some counterintuitive results. Observe that the Sider and Braun’s motivation in developing their theory is to tweak supervaluationism to eliminate its problems. If their theory entails consequences that are more counterintuitive than its supervaluationist counterpart, and I believe I have shown that it does, then his motivations are undermined and no real progress has been made.

Furthermore, the supervaluationist is able to give a reply to (4) that is analogous to the Sider and Braun’s, and apparently without the strange consequences that they are forced to accept. Keefe writes, ‘It is not always the case that [(4)] would be assented to: when someone is aware of the slippery slope down which the premise may lead us, a reasonable and common response would be to refuse to endorse that premise. In general, test our intuitions about p in contexts where the inferential consequences of p are salient, and the intuitions can be different from outside specific contexts of that kind’ (183-184). Thus, Sider and Braun’s reply to (4) says nothing that supervaluationism cannot also say. Their modifications to supervaluationism bring nothing new to the table regarding this issue.

V. Approximate Validity and the A Operator

So far, Sider and Braun have been able to retain classical logic to a great extent by giving a non-classical semantics for the notion of approximate truth, much like the supervaluationists do for truth simpliciter. To go along with this account, they need to provide a definition for validity in natural languages. They do not want to say that we cannot draw any inferences in a vague language, and so they must explain how the notion of validity applies to such a language. An argument is approximately valid (or, as they confusingly call it, ‘valid simpliciter’) iff every admissible precisification that makes its premises true also makes its conclusion true. The definition is analogous to Williamson’s definition of local validity

\^{10} (147)^{11}.
Williamson (148) and Keefe (174) agree that local validity should not be the supervaluationist’s conception of validity because of their identification of truth with supertruth. Although Sider and Braun do not make that identification, it is not clear that they are able to avoid related worries about their own theory. They claim that vague sentences which we intuitively want to say are true are usually merely approximately true (true on all admissible precisifications). The tenor of their argument is that our ordinary practices are perfectly acceptable, as long as we ignore or idealize away the vagueness. But on that view, because they thinks of approximate truth as ‘truth, vagueness ignored,’ approximate validity should be the necessary preservation of approximate truth. Yet, his definition of approximate validity does not involve the notion of approximate truth. Thus, to be consistent with his overall project, approximate validity should be defined thus: an argument is approximately valid just in case, necessarily if the premises are approximately true (supertrue), then the conclusion is approximately true (supertrue)\(^{12}\). This is an analogue to the standard supervaluationist definition of validity (global validity, in Williamson’s terminology.) In fairness, Sider and Braun point out that on their definition, if an argument is approximately valid then, trivially, it necessarily preserves approximate truth. What they, importantly, leave out is that the converse fails – approximately valid arguments, in their sense, is merely a (proper) subset of arguments necessarily preserving approximate truth. Consider the inference from ‘\(\varphi\)’ to ‘\(\varphi\) is approximately true’, for some vague sentence \(\varphi\) involving a borderline case. Clearly this inference preserves approximate truth: if \(\varphi\) is approximately true, then it is approximately true that \(\varphi\) is approximately true. However, since \(\varphi\) is borderline, there will be some precisifications on which \(\varphi\) is true, and others on which it is false. So, there will be precisifications on which ‘\(\varphi\)’ will be true and ‘\(\varphi\) is approximately true’ will be false – the argument is not approximately valid in the Sider/Braun sense, thus showing that the preservation of approximate truth and the Sider/Braun definition of approximate validity come apart.

On the conception of approximate validity that I have argued Sider and Braun have forced themselves into, approximate validity preserves all valid inferences and rejects all invalid inferences for a language containing only the standard connectives \{&, ∨, ¬, →, ≡\}. Keefe’s arguments to this effect for supervaluationism (175-176) show that the claim holds for the Sider/Braun theory as well. However, the proposed account runs into trouble when a unary ‘A’ connective for approximate truth is introduced, where ‘\(A\varphi\)’ is true on a precisification if and only if \(\varphi\) is true on all precisifications. They might claim that the ‘A’ operator is somehow inappropriate in his system,
but then he loses one of the main motivations behind the supervaluationist framework: it gives a semantics which is able to express ‘φ is a borderline case’ and ‘φ is a clear case.’ I won’t delve into the technical details, but it turns out that Sider runs into all the same problems with ‘A’ that supervaluationism does with the ‘D’ operator. In other words, the following are all true in his theory:

\[
\begin{align*}
\phi & \models A\phi \\
\neg A\phi & \not\models \neg \phi \\
\not\models \phi & \rightarrow A\phi
\end{align*}
\]

…and so on (see Williamson 151-152).

Contraposition, the deduction theorem, proof by cases, and *reductio ad absurdum* fail on both the supervaluationist account and the Sider/Braun theory, when the deduction involves the ‘A’ or ‘D’ operator. Of course, Keefe’s reply to this objection (178-181) can be adapted to suit the new theory. On this issue, the Sider and Braun are no better off than supervaluationism.

**VI. The T- and F- Schemas**

Sider and Braun argue that they are able to retain the disquotational truth predicate and all of the associated T- and F-schemas (“‘φ’ is true iff φ”, etc). They argue this point at length in their paper; and I will not reproduce the entire argument here. The idea is that their theory does not modify the classical notion of truth, it merely limits the notion’s scope. Thus, the T-sentences remain true, but there are a lot less of them. For example, the T-sentence “‘Johnk is tallj’ is true iff Johnk is tallj”, where k and j are legitimate precisifications of the vague terms, comes out true. However, there are strengthened versions of supervaluationism which are able to accommodate for a disquotational truth predicate – Sider and Braun makes no advances on this point beyond any but the most simplistic versions of supervaluationism. Furthermore, the theory runs afoul with instances of the T-schema that contain vague words.

(5)“‘Snow is white’ is true iff snow is white.”

Most English speakers would be willing to assent to (5), and therefore, by their norm of assertion, it had better be approximately true. Indeed Sider and Braun write, “it would be unwelcome if accepting the present theory
required radical changed in linguistic behavior such as ceasing to assert sentence like (5).” They claim that (5) “is certainly approximately true.” However, the certainty they seem to notice is mysterious. ‘Snow is white’ is the name of a vague sentence, therefore it does not, strictly speaking, name a true sentence. So, on each precisification, the left-hand side of (5) is false, so “snow is white” is approximately false. The right-hand side of (5), however, is true on some precisifications and false on others – it is neither approximately true nor approximately false. Therefore, the biconditional in (5) is not approximately true; therefore (5) is unassertable. By the authors’ own admission, their theory is unwelcome.

VII. Conclusion

The motivation behind Sider and Braun’s account of vagueness is to “preserve what is right about supervaluationism while discarding the problematic baggage” (16). If the theory fails to achieve this goal, then their project has failed. They claim that vague sentences lack truth values, which they support with three arguments: an analogy to physics, another to ambiguity, and a claim about the nature of propositions. I believe that I have cast doubt on the strength of all three arguments. Of course, there may be other arguments which justify the claim that the notion of truth only makes sense relative to a precisification. Perhaps those arguments are more successful. If so, then Sider and Braun might simply take them on and leave the rest of their theory intact. On that assumption, how successful is their theory in solving the supervaluationist’s difficulties? They seem to be able to retain a disquotational truth predicate for perfectly precise sentences, which some supervaluationists cannot. Yet, the theory has trouble with T-schemas containing vague words. I have argued that their account of ∃ and ∨ is no better, and is in fact worse, than the supervaluationist approach. Furthermore, their theory is apparently unable to resolve the logical anomalies involving the rejection of certain fundamental rules of deductive inference that face the supervaluationist. Similarly, as soon as an ‘A’ operator is introduced, Williamson’s arguments against the supervaluationist account of higher order vagueness can be adapted to undercut the proposed theory. Thus, it is my position that the Sider/Braun theory has been built on questionable grounds, and that it does not amount to a significant advancement over supervaluationism. Another approach is needed†.
NOTES

1 Ted Sider (Rutgers University) and David Braun (University of Rochester), ‘Vague, So Untrue’ (http://fas-philosophy.rutgers.edu/~sider/papers/vagueness.pdf).

2 Throughout this paper, I will use the presentation of supervaluationism in chapters 7 and 8 of Rosanna Keefe’s *Theories of Vagueness* (Cambridge: Cambridge University Press, 2000) as the default supervaluationist semantics against which I shall compare Sider and Braun’s theory.

3 This argument played a more prominent role in an earlier version of Sider and Braun’s paper. In ‘Vague, So Untrue’, much of the argument has been removed or hidden. There are, however, several passages where it is clear that the analogy to physics still lurks in the background, so I include my objection to it here.

4 Russell in ‘Vagueness’ (1923) in Rosanna Keefe and Peter Smith (eds.) *Vagueness: A Reader* (Cambridge, Mass: MIT Press, 1997) makes a similar argument that logic assumes precision, and thus does not apply at all to vague languages.

5 His argument is similar to (and, unfortunately, just as brief as) the argument in Kit Fine, ‘Vagueness, Truth, and Logic’ (1975) in Rosanna Keefe and Peter Smith (eds.) *Vagueness: A Reader* (Cambridge, Mass: MIT Press, 1997) (135-136). Sider departs from Fine, however, by arguing that the supervaluationist approach fails for the ambiguous and vague sentences.

6 Although, see David Lewis, ‘Logic for Equivocators’ (*Noûs*, Vol. 16, No. 3. (Sep., 1982), pp. 431-441) for alternative ideas on how this might be done in a non-classical logic.

7 Perhaps the analogy could be restored by accepting a Quinean indeterminacy of reference thesis and arguing that ambiguity actually is a case of indeterminacy. But on that conception, ambiguous words are still no more indeterminate than everyday words, and it’s not clear that the argument could be preserved.

8 This objection is similar Williamson’s objection to global nihilism (169).

9 It is plausible that there are others, e.g., be relevant, don’t give too many details, etc. See, for example, Paul Grice, ‘Logic and Conversation’ in *Studies in the Ways of Words* (Cambridge, Mass: Harvard University Press, 1989).

10 As MacFarlane points out, this notion may be formally defined as: $\varepsilon \forall \nu (\forall \alpha \varepsilon \nu A \text{ is true on } \nu \rightarrow C \text{ is true on } \nu)$, where $\nu$ is an admissible precisification, $\Gamma$ is the set of premises $\{A_1, ..., A_n\}$, and $C$ is the conclusion.

Again, due to MacFarlane: $\mathcal{E}((\forall v \forall_{AE} A \text{ is true on } v) \rightarrow (\forall v C \text{ is true on } v))$.

†This paper was originally presented to John MacFarlane’s vagueness seminar in Spring 2002 at UC Berkeley. I am grateful to MacFarlane, Geoff Georgi, and George Watson for their comments on the initial paper.
“Not falling, not ignoring
A pair of Mandarin ducks
Alighting, bobbing, anywhere.”

~ Nan-O-Myo
Induction is the rational cognitive process whereby we infer novel conclusions about states of affairs in the world based upon information that we already have. An inductive conclusion could be about the future or the present, and the information contained within the premises could be about the present or the past. Here are two examples:

I. Inductive reasoning about some present circumstance of which we do not yet know:

Regarding *Stycast Epoxy*¹

P(1) The can that I am currently holding in my hand, from which the epoxy came states that it cures² at room temperature.

P(2) Last time I poured it in a mold it cured in about a day at room temperature.

P(3) This time when I made the epoxy mold, I did it in the same way as last time.

P(4) I made the epoxy mold yesterday, and it has sat for a day at room temperature.

C (1) The epoxy that I am currently looking at is cured.

II. Inductive reasoning about some future circumstance:

Regarding whether or not a carrier pigeon will return to its coop when it has finished its delivery:

P(1) I have often in the past given delivered messages with this pigeon, and it has always returned to its coop when finished.

C(1) It will, when finished with its delivery, return to its coop.

Notice the common feature in inductive reasoning, that it involves some sort of generalization upon the facts as they have already presented themselves, and then making a prediction about some particular circumstance(s) in the world of which we have no factual knowledge. This is done regularly
in our mundane existence as a matter of commonsense. Science as an extension of commonsense, although a much enhanced version, relies heavily on this sort of reasoning. For science takes the observable data from the world and constructs hypotheses and theories as to why and how the observable phenomena behaves as they do; furthermore, science also formulates predictions based upon these generalizations (hypotheses and theories). The pervasive nature of induction and its usefulness make it a worthwhile and fascinating topic for discussion.

However, in discussing induction philosophers have discovered that certain serious issues present themselves; these must be addressed. It is hoped in so doing that we would improve upon our practice of inductive inference making, much like the study of soil science improves upon the practice of farming. The focus of this paper will be twofold. First, it will discuss the traditional problem of induction and its proffered solution. Second, it will discuss the new problem induction, its proffered solution, and its merit. So to get started, let us consider a parable Parables are useful because they help to give a good vantage point from which to see the issues to which we must address:

He never cared for these new electric cutting tools. “The blade moves too fast”, he thought. “I just might lose my head if the damn thing flies off”, he said aloud to himself. Turning around for the pen entrance, answering half to the chickens and half to himself, “I won’t be losing my head today. Even if the thing flies off, the guard will protect me; keep the thing from flying up into me face.” The chickens, seeing Luc, run from all directions towards him, eager, hungry. It did not help that he had the all too familiar red feeding bucket. A flutter of pity surged through Luc’s heart as he saw them, so certain of food, yet so naïve to the dinner plans of others. He paused a moment, “those dumb beast. Don’t they see the axe in my right hand? Ah, it’s just as well…. I just wish I didn’t have to be the one to hand them the bad news….”

The above parable demonstrates the two issues which concern us in this paper. The first is the traditional problem of providing a rational justification of induction (e.g., inductive inferences like those used by the gentleman in the first paragraph of the parable above). The second problem that will be discussed in this paper is the new problem of induction: namely, how to de-
termine which regularities are lawlike and which are non-lawlike. The poor chickens in the second paragraph of our parable obviously made a mistake in this all too important task. However, we should be able to do a lot better then they did. So, let us begin with the traditional problem of induction.

The Old Riddle of Induction

The traditional problem of induction was that of justifying it rationally; thus, it was incumbent upon any philosopher whose focus was logic or epistemology, who cared about the problem of induction, to provide some way to do this. It was certainly not any easy task, for to rationally justify it would seem to require logic. The two accepted logical methods of argumentation that readily come to mind that may show some promise are deductive or inductive. However, both of these are problematic in themselves for reasons forthcoming.

Before we go in to the problems of deductive and inductive argumentation that would attempt to rationally justify induction, it is necessary that we first stipulate an explicit requirement for rational justification of induction from which to start, something fairly uncontroversial. The requirement that seems best suited is the following:

Requirement for Rational Justification

Induction or an inductive system is rationally justified if and only if the arguments of which it assigns high inductive probability yield true conclusions from true premises most of the time. Furthermore, the degree of inductive probability must directly correspond to the frequency with which these arguments that use induction or are part of some inductive system produce true conclusions from true premises most of the time, higher inductive probability for inductive arguments that yield true conclusions from true premises most of the time and lower inductive probability for arguments that yield false conclusions from true premises most of the time. By “true,” I mean that the information contained within these premises is factually correct about either some feature of the natural world of convention (Logic or Mathematics).

The notion of inductive probability needs to be explained, for it is assumed in the formal characterization of the requirement for rational justifica-
tion of induction. Inductive probability is the numerical value assigned to an inductive argument that guarantees true conclusions from true premises most of the time. The numerical value ranges between 0 and 1 (i.e., (0-1]). Here are two examples of inductive arguments that have been assigned inductive probabilities of differing magnitudes:

**Argument with High Inductive Probability:**
1) Every winter (in the past) ducks migrated to my home town.
2) Therefore, this winter they will migrate to my home town.

This argument would have an inductive probability of 0.85 or 85%; whereas, the following argument would have much less.

**Argument with Low Inductive Probability:**
1) On all birthdays that I have ever had I have been less then 26 years old.
2) Therefore, on my next birthday I will be less then 26 years old.

This argument would have an inductive probability of 0.25 or 25% for obvious reasons. The reason that this argument would have low inductive probability is that it assumes the future will be like the past, but not in the relevant respect. (this will be discussed in depth shortly.) With the requirement and illustrations of it in mind, let us now examine, each in it turn, the deductive and inductive justification of induction.

**Deduction**

If we try to justify induction deductively then what we must use are deductively valid arguments. The premises in a deductively valid argument yield no novel conclusions. In fact, everything contained explicitly in the conclusion of a deductively valid argument, are implicitly contained in the premises. For example, consider the following argument:

P(1) All men are mortal.
P(2) Socrates is a man.
C(1) Therefore Socrates is mortal.

Thinking set-theoretically, consider Socrates as being in element in the set “Men.” Consider another set “Mortal”, that subsumes (i.e., the set “Men”
in included in the set “Mortal”) the set “Men”. The set “Men” would be a subset of the set “Mortal”, and the element “Socrates” would be an element of this set too. Therefore, it could be demonstrated in set-theoretic fashion that Socrates is mortal. As can be seen however, the conjunction of premise one and two state this implicitly, whereas the conclusion states the same information explicitly; hence, the conclusion states nothing over and above the premises. The conjunctive premises and single conclusion can be thought of as two different converging descriptions of the same phenomenon – Socrates’ mortality.

If the conclusion of a deductive argument states nothing novel, nothing over and above the premises, then it states what we already know, from the past and present. It states nothing about the future, of which we have no knowledge. Yet, again as would be readily agreed to, induction and inductive arguments are supposed to yield novel data, facts not hitherto contained in the premises of the selfsame argument. So, it seems that a deductive strategy would be inadequate.

**Induction**

What about using some kind of inductive procedure to justify inductive practice? As will be shown, this strategy doesn’t look any more promising; consider the following type of formal inductive argument that might be given to justify induction:

(1) Arguments that have been judged by some system of induction or other to have high inductive probability have yielded true conclusions from true premises most of the time in the past.

(2) Therefore, such arguments will yield true conclusions from true premises most of the time in either the past, present or future.

Notice though that this just begs the question, for to justify induction by an inductive argument is just to use the reasoning in question to justify itself, this would be assuming the very thing we seek to justify, in order to justify. This is not only logically unsound, but also this strategy is devoid of any substantive answer to the original question, and metaphorically speaking, would be just as useful as using an unbroken knife blade to cut itself. So, an inductive justification of induction would be useless.
**The Uniformity of Nature**

The task of justifying induction can be viewed from a different perspective though, from that of the principle of the uniformity of nature; this principle states that events that occur in the natural world do so regularly, in the past, present and future, and that these regularities are causally determined. For example, all objects accelerate towards the center of the Earth at roughly $10\text{m/sec}^2$ due to gravitational attraction between any massive body in question and the Earth (understood as a very massive body). Consider also two inductive arguments that assume this principle, so to get a better understanding of it.

If you were accustomed to eating the entire strawberry, head and all, when eating strawberries, and a friend comes along and sees you doing so, and says to you that “eating the head of a strawberry would give you a stomach ache”; you could respond:

*Inductive argument 1:*
1) I have often eaten the entire strawberry in the past, and I have not had a stomach ache that could be directly traceable to having done so.
2) Therefore, eating the head of this strawberry will not give me a stomach ache.

While this argument would be the commonsense reply, consider using the same type of argument in a more rigorous setting. Suppose there’s a woman named Marguerite who is an astrophysicist working on a team of people that construct deep space probes. She is being interviewed by a reporter about a probe that was recently sent out to an area of space that we can not directly observe, but know is there. The reporter asks what assurance they have that the probe will function comparably to directly observable situations. Marguerite responds with the following inductive argument:

*Inductive argument 2:*
1) Principles Q, R and P have correctly described and predicted behavior in many circumstances that we can observe.
2) Therefore Principles Q, R and P can correctly predict the behavior of the probe in many circumstances that we have as of yet not observed or can’t observe. (Again, her reply would seem sufficient).
Notice that the common assumption supporting these two arguments is that future events in the world will be just as they have been observed to be in the past; that in some way nature is uniform, in the way explained above—that there are a myriad of causal relations that occur regularly. Nature certainly is uniform, but not in all respects; and we do not as of yet have a system or at least a set of definitions that will tell us in what respects nature is/will be uniform. Of course we eventually figure out some of the ways in which nature is uniform, but it is formally done through experimentation, and from amassing data, concerning what we have or presently are observing, and not from what we have yet to observe. Without question begging, we have no system or set of definitions that will tell us ahead of time in what respects nature will be uniform. Even in cases where we have made many observations of events in some circumstances, and get some idea that the events will take place uniformly, we could be wrong—dead wrong, as was shown in the slaughter-house parable earlier. Consider:

Imagine a man working in some university bio-physics lab. He regularly has observed that a low-ground-level fog comes out of the adjacent room at 2pm everyday. The substance that he observes is just evaporating liquid nitrogen. Today, he is in the lab and makes the same observation, yet this time the gas is some sort of neurotoxin. He drops the ground and begins having a seizure.

The above example should make the fallibility of our beliefs salient: we do not know enough about how nature is uniform to understand and predict how it will be uniform.

So what to do? We can try to establish, through argumentation some system that will give the conclusion that nature in uniform in X, Y, Z… respects. What sort of reasoning though would be able to establish such a conclusion concerning some system? In order for the reasoning to be cogent, it must be either deductively valid or inductively strong. Deductive arguments do not give any novel data so they cannot suffice to establish such a conclusion. We are then left with inductive arguments. However, if we put faith in inductive arguments then it is incumbent upon us to answer the question “why should you trust these arguments?” It does no good to say “because nature in uniform,” for this would again beg the question. We aim to rationally justify induction, but deductive arguments fall short. Inductive arguments end up being like so many dogs – just chasing their tales. The principle of the uniformity of nature cannot help us, for we have
no knowledge, or any way of knowing, in what respects nature is/will be uniform. It seems that we are at an impasse.

*The Turn Towards Practice*

Let us now turn then to Hume’s way out of this riddle. Hume, after surveying the landscape he saw many of the pitfalls that have been discussed in some detail thus far in this paper. So, he decided to look in the most unlikely place imaginable, for philosophers of his day and for some since – behavior. It was to inductive behavior that Hume turned, not for the origins of inductive reasoning, but for its rational justification\(^7\). In order to understand the main thrust of his strategy, let us consider how deductive inferences are justified. These sorts of inferences are justified by showing that they are well supported by rules, but how are the rules justified? The following parable will demonstrate how.

Thoth is an Egyptian philosopher and ascetic whose aim it is to codify some of the accepted and agreed upon deductive practices of his day. These practices have been unanimously used for some time, but have not heretofore been codified. So, he constructs some formal rules that formally support the deductive practices. Future generations can then refer to these rules when disputes of deductive practice arise. The rules themselves, however, are justified by the sound deductive practices. *Thus, the practices are demonstrated and justified by the rules, and the rules are justified by the practices*, both then could be thought of as justified by being brought to agree with each other.

What is true in the parable above for deductive inference practices and their corresponding rules is likewise true of inductive inferences and any rules of a system of induction, *that inductive rules are justified by viable inductive practices and these selfsame rules justify inductive practices*. So, the old riddle of induction has been reasonably solved\(^8\).

*The New Riddle of Induction*

With the old riddle of Induction solved one might think that it was high time to rest easy. Yet, problems remain, and they do so because, although
Hume was correct in turning his focus on inductive practices, he was not thorough enough. Roughly, he mentioned that regularities observed give rise to habitual expectations, and predictions based on these regularities are normal, even valid. However, he failed to address the fact that some regularities do not establish habits of expectation, whereas others do (See Encyclopedia of Philosophy Vol. 3-4). What accounts for this, and how are we supposed to distinguish between regularities that establish these habits and those that do not? These two questions can really be conjoined into one: How do we distinguish between lawlike and non-lawlike regularities?

Projectibility

Before we get into Goodman’s solution to the aforementioned question unanswered by Hume, we need to take a brief detour to get a better handle on fundamental concepts that will be used to give Goodman’s solution—projectibility.

Projectibility is a methodological activity whereby one takes a predicate, which is a formal representation of the habits of expectation that correspond to some set of observed regularities, and then makes some prediction about the future based upon the nature of such a predicate (i.e., the nature of the habits of expectation that correspond to some set of observed regularities). Consider the complex predicate M. The predicate “M” ranges over the expectations that metal will expand when heated. This expectation corresponds to the observed regularity that metal does indeed expand when exposed to some heat source. Predicate “M” is projectible insofar as we can make viable predictions that the next piece of metal that is exposed to some heat source will expand as well. With the above characterization of projectibility, we can move to the discussion of Goodman’s solution to the new riddle of induction.

Entrenchment

“Entrenchment” is a relational term. It can be characterized as a dyadic predicate (i.e., Exy {“x” is better entrenched than “y”}). Thus, Goodman’s answer to the question of lawlike regularities can best be formulated by the following:

A predicate “P” is better entrenched than “Q” if “P” and all predicates co-extensive with it have in fact been projected more often then the predicate “Q”
We can now consider two pairs of predicates that will help clarify the above formal characterization. Consider the predicate P. “P” ranges over, or references the linguistic concept of arachnid, which of course has reference to arachnids existing in actual fact (i.e., “P” \{arachnid\}). The predicate “Q” ranges over arachnids of a certain time period and insects of still another time period or, “Q” ranges over arachnid from some indefinite past to time \(t_1\) and insect from time \(t_2\) to some indefinite future (i.e., “Q” \{arachnid/insect\}: \((\infty -t_1], [t_2- \infty)\) ). Take also the predicate “R” which ranges over the linguistic concept “square,” in like manner to the predicate “P,” references actually existing square-shaped objects. Consider also the predicate “S” which ranges over all squares from, again, some indefinite past to time \(t_1\) and triangles from time \(t_2\) to some indefinite future (i.e., “S” \{squares/triangles\}: \((\infty -t_1], [t_2- \infty)\) ). “P” and “R” are better entrenched predicates than “Q” and “S,” because “P” and “R” have been used more frequently in making viable predictions than “Q” and “S”; thus, habits of classification have been established by which predictions have been habitually made. These predictions and classifications are not just habits, but they are classifications and predictions that are pragmatically viable, and have been formed into methodological habits. The pragmatic viability of these classifications and predictions consist in that they consistently produce results that are useful to us, in terms of science, technology, civil management, etc. For just think if the habits of classification and corresponding prediction were adopted such that predicate “S” was in use, then such a discipline like civil engineering or architecture would be adversely affected, for obvious reasons.

Even if it is the case that two predicates are equally supported by evidence, one predicate can still be better entrenched than another, and that better entrenched predicate is to be preferred. Reason being, it is tried and true and has worked in the past many times, and promises to work in the future. The notion of entrenchment applies equally well to a composite of predicates. A theory or hypothesis just is a composite of predicates, so the notion of entrenchment would apply equally to a theory or hypothesis. For example, consider two theories about consciousness and its relationship to the brain. The first theory (T1) states that consciousness is caused by neurophysiologic processes. It does not exist independent of some nervous system or other, but is entirely dependent on it. The second theory (T2) states that consciousness is some form of energy that exists entirely independently from neural matter, and that it is present as a neurobiological fact in direct proportion to the complexity of the nervous system, the more complex the
nervous system, the more consciousness as energy is present.

T1 is the hegemonic view of today. It has been established for some time. Classifications and predictions made in light of this view have been robust and consistent for 85% of the time; furthermore, this view has often been projected in the sense that is meant in the paper. T2 has not been around the block, so to speak nearly as much as T1, nor has it produced classifications and predictions that are robust and consistent; thus T1 is better entrenched then T2, though they are evidentially equivalent. By evidentially equivalent, I mean that the theories are underdetermined by the observable evidence (i.e., that the observations made by scientists studying consciousness and the brain equally support both theories). However, T1 is accepted despite it being underdetermined by the observable evidence, because it is a better entrenched composite of predicates. As mentioned above, the notion of entrenchment as formally characterized is a relational term, logically speaking. Since it is relational, it admits of degrees—some predicate is better entrenched then another predicate, it being entrenched to a greater degree. So, what then would be the greatest degree of entrenchment, or what would be an ideally entrenched predicate?

Ideally entrenched predicates are natural kinds. Natural kind predicates are those that symbolize highly consistent habits of expectation, which correspond to highly consistent regularities. Such predicates symbolize the following regularly occurring phenomena: motion, heat, electromagnetism, triangles, mammals, etc. In fact, the highly consistent habits of expectation that are formed by these regularly occurring phenomena are the reason why these predicates are ideally entrenched. How are these sorts of predicates identified? By the fact that these classifications and predictions made on them have proved successful in the past, and give promise of being successful in the future qua their reliability in the past.

Goodman’s solution seems, prima facie to be a good one. However, there are some questions that arise about it, questions that need to be addressed. The merit of his view need be accessed, to that we next turn.

Assessment

This section will discuss two merits of Goodman’s view: conservatism and reasonableness. One merit of such a view is that it is methodologically conservative. It is such by relying on that which has been tried and shown to work in the past, only predicates that have been robust classifications and stable bases for predictions are those that are part of the canon of acceptable predicates. This is a merit because it sets up a filter that bars entry of any
spurious and unfounded predicates (e.g., those that would lead to some sort of contradiction, or not symbolize any true regularity of the natural world).

Another merit of his view is that it appeals to common sense. It does so by using the reasoning that is part and parcel of our lives – inductive reasoning. Cognitively, we are so accustomed to using this form of reasoning that it is easy to see Goodman’s point. This is definitely a strong point for his view, for if the investigations of philosophy can clarify our common sense views and yet not throw them out, so much the better for it. Though, the merits of Goodman’s view do not in any way detract from the fact that there are some concerns that his view brings up, that again, need to be discussed.

What is the relationship between entrenched predicates and the discovery of new phenomena, new regularities? Furthermore, how are new predicates incorporated into the canon of entrenched predicates? Let us begin with the second question: how are new predicates incorporated into the canon of entrenched predicates?

There are two ways in which new predicates can be incorporated into the canon of acceptable predicates. The first is that a composite of well entrenched predicates that have been used to formally characterize the practice of using fossil fuel, must range over similar phenomena with respect to the composite of predicates used to discuss hydrogen fuel cells, such as energy transduction, heat, fuel, etc. This ranging over similar phenomena can be though geometrically as the similar phenomena being in the intersection of the predicates. This sharing of ranges allows, methodologically speaking, a smooth transition from formalization of past observations, that lead to the discovery of the fossil fuel driven combustion engine, to the formalization of the present observations concerning the experiments currently underway to develop hydrogen fuel cells. The key saving methodological piece is that of similarity of elements within the intersection of the sets (composites of predicates).

The second way in which new predicates can be incorporated into the canon is that of an isomorphism between the formalizations themselves, and between their predicate structures. For example, consider the classic debate between the heliocentric and geocentric conceptions of the solar system. There were structural similarities between the two views, such as eccentricity of orbits. However, despite the structural similarities, the geocentric view just accounted for more of the observational data that was able to be detected at the time. It was not only its better capacity to account for observations that made the heliocentric view better, but that it was methodologically more parsimonious, elegant, robust, etc. Thus, the heliocentric view was accepted and the geocentric view dropped. Generally speaking,
the point is that new predicates are adopted into the canon of acceptable or well entrenched predicates insofar as there is some structural similarity between the new constructed predicates and the old predicates, but the new predicates has structural features that make it much more appealing, such as elegance, parsimony, greater, precision, more robust, etc. Thus, there are two ways in which predicates can be brought into the fold of well entrenched predicates: First, there has to be some similarity in the phenomena that is referenced or ranged over by the new and old predicates. Second, there has to be structural similarities in the predicates themselves, yet the new predicate must be more methodological advantageous for reasons already mentioned, and the new predicate must be able to better account for the same or similar observational data as the old.

The first question (what is the relationship between entrenched predicates and the discovery of new phenomena, new regularities?) can now be answered. The relationship between well entrenched predicates and the new phenomena is similarity. There is a limit to how far similarity can take us though. It cannot bring us in touch with truly novel discoveries. These discoveries may be so new as to put considerable strain on our accepted scientific methodology, of which well-entrenched predicates are an integral part. With well-entrenched predicates we can really only move from the known to the known.

**Reflections**

The old riddle of Induction was concerned with giving a rational justification of it. We discussed that to do so by means of deductive logic is inadequate; furthermore, to do so by means of inductive logic was circular. The solution then was to turn to inductive practices. An examination of the practices would give the sought after rational justification.

The new riddle of induction was concerned with finding a way to isolate viable regularities in the world that would form viable habits of expectation. The way to do this that was proposed was to look again at our practices, to see which predicates have been viable in the past so as to get an idea of which ones actually classify viable regularities of the natural world, and upon which predicates are suitable bases for predictions. In fact, in most cases, these are the same predicates. We also discussed briefly the merits of this solution, and then some its problems. What can we glean from our discussion?

The strategy for the solutions to the old and the new riddle of induction
are similar in that they both turn towards actual methodological practices as a basis of their respective proffered solutions. Also, regarding the solution to the new riddle of induction, the issue of distinguishing lawlike from non-lawlike regularities is just another form or instance of the general problem of developing a way to know ahead of time in what ways the natural world will be uniform—the problem concerning the uniformity of nature. This only stands to reason because the issue that was not addressed in the solution to the old riddle of Induction was that of the uniformity of nature.

WORKS CITED


NOTES

1. *Stycast Epoxy* is an adhesive resin. It consists of two parts, two different liquid chemical substances. It is used to hold parts together that have not been bonded mechanically. For example, it is just in the fabrication of particle detector housing towers for the *Cryogenic Dark Matter Search* group in astrophysics at the University of California, Berkeley.

2. By “cures”, I mean that it solidifies.


4. I am using convention in the Quinian sense. Please see *Word and Object* and the article “Truth By Convention”, W.V. Quine.

5. This paper is not an attempt at an explanation of deductive validity. In fact, a definition of it is assumed. However, if a reader wishes to know more about the notion of “deductive validity” I encourage her to peruse through *The Logic Book*, by Merrie Bergmann.

6. It must be admitted here that induction need not solely be about the future. It can be about the present as well. The point is that it must yield novel conclusions, information that is not contained implicitly in the premises of the argument. Speaking in temporal emphasis on the future just drives the point home more clearly.

7. I do not here wish to get into whether or not Hume indeed did pull this off, which is a whole other paper in itself, but just the main thrust of his point.

8. There are those that think that the problem is still running amok leaving havoc in its wake. This paper is not concerned with such thinkers. Though, a fascinating and fruitful discussion about these thinkers’ views can be had.

9. For more on dyadic predicates, of the form $SExy$, see “The Seesaw-Effect and Systemic Insight in Couple Therapy” http://www.noga-nabarro.com/
The phrase “habits of classification” refers to the regular practice of classifying in a certain manner. For example, calling a flat upright piece of wood or metal on hinges a door.

Since, the traditional problem of induction has been solved, which Goodman agrees with, it is not questionable to use basic inductive strategies as a basis for his solution, induction being rationally justified.

I say this because a view that has been long-established yet clarified is much easier to accept, even though the clarifications may be novel; then a the total replacement of a long-established view, which due to the neurobiological capacity for habit formation would be much harder to accept.
PAOLO SANTORIO

Two Externalist Theories of Perception?

“From its *seeming* to be so - or to everyone - to be so, it does not follow that it *is* so. What we can ask is whether it can make sense to doubt it.”

~ L. Wittgenstein, *On Certainty*
In this paper I aim to draw a comparison between Searle’s and the disjunctive theories of perception. In particular, I will try to contrast the cases of perception and hallucination, and the way the two theories provide an explanation for them. What I hope to show is that, although one would expect to come to quite different solutions, there is an unexpected and in some way striking similarity between Searle’s account of perception in terms of Intentional content and a radically externalist account as the disjunctive theory.

Traditional Intentional theories of perception assume that, both in cases of perception and cases of non-perception (like illusion or hallucination), the subject has a visual experience of the same kind. The disjunctive theory is grounded just on the denial of this “higher common factor.” According to the disjunctivist, either I perceive an object or a state of affairs, or I am subjected to an illusion. But while in the latter case what is presented to my mind is evidently not something that is in the world, in the former “the object of experience does not fall short of the fact: the object of experience is the fact.” In other words, the idea of an experience which in some way stands between the mind and the world is abolished, and what is presented to the subject is the fact itself.

Searle’s account of perception is in terms of Intentionality. When I have a perception, e.g. when I see something, I have a visual experience that has as its Intentional content the state of affairs that is perceived. The content of the visual experience is always propositional in form, namely an experience is not of an object, but that such-and-such is the case. Searle’s theory diverges from other representational theories of perception, as he claims that what is seen is not the visual experience itself, but the state of affairs in the world—the fact. I see a certain state of affairs, and I have a visual experience of that state of affairs. By contrast, when I am subjected to an illusion, I do have a visual experience, but there is nothing in the world corresponding to it; as a matter of fact, I do not see anything, and I only have an impression of seeing. The conditions of satisfaction of the visual experience are self-referential in the sense that they specify both that such-and-such must be the case, and that such-and-such being the case must cause the visual experience. Hence, if I see the ocean, the conditions of satisfaction of my visual experience are

I have a visual experience (that there is the ocean in front of me, and that there is the ocean in front of me is causing this visual experience).
This kind of self-referentiality is needed, as it is not sufficient that the mental content of the perceiver just matches the state of affairs in the world; there must also be some kind of relation between the two so that the former would not be what it is were it not for the latter. Self-referentiality, and the concept of Intentional causation, can thus be used to deal with counterexamples of the Twin Earth type. What makes my perception different from the type-identical perception of my twin on Twin Earth is the fact that the relation of causation is part of the conditions of satisfaction of the visual experience. In this way the two perceptions come to be different, as the content of the two visual experiences specifies two different objects.

I want now to use the two theories to analyze a particular case of hallucination. Let us imagine that I am on the beach, facing the ocean. At the same time, I am the subject of a psychological experiment, so that my brain is directly connected to a quite sophisticated electronic machine. When the machine is not activated, I ordinarily perceive the sea. But when an experimenter pushes a certain button on a control panel, the connection between my brain and my optic nerve is suspended, and the perception is substituted by an image that is obtained by direct stimulation of my brain cortex. Now, the image that is projected in this way into my brain comes from a micro camera that lies on my face, in the exact middle of my eyes. The result of all this rather complicated apparatus is that, whether I perceive the sea with my own eyes, or I simply receive an electric stimulation in my brain, I have exactly the same phenomenological experience, and there is no way I could distinguish the perception from the image induced artificially. We could also imagine that the experimenter pushes and releases the button continuously, so that I pass from a state of perception to a state of illusion, though my experience is an uninterrupted stream, and I do not in any way notice the change. Now I assume that, although I come in some way to have a veridical image of the world, nobody would want to say that the experience generated by brain stimulation is a case of perception. When the experimenter pushes the button, my perception ceases, and I have a particular kind of illusion, which, for contingent reasons, happens to portray the real world. Insofar as perception is a phenomenon involving sensorial experience, and not just the presence in my brain of data matching the world, the present case is not a case of perception.

If we want to look at the issue in causal terms, the cause of my perception is not simply that there is a such-and-such state of affairs in the world. Rather, the content of my hallucination is the outcome of a causal chain including all the machinery, the fact that the experimenter pushes the button, and, just among the other factors, the image that is captured by
the camera. So, it seems clear that even proponents of causal theories of perception would not want to say that in the case of brain stimulation I do effectively perceive the ocean.

Now, what is the analysis of the case just exposed in the terms of the two theories? Although at first sight we would think to obtain completely different results, I believe that, this example could make us discover an interesting similarity between the two accounts. The analysis in terms of the disjunctive theory is fairly simple. At each point in time, I am either having a perception, or I am having a hallucination. The fact that there is a phenomenological continuity, which does not allow me to distinguish when one condition ends and when another begins, is completely irrelevant. There is not a common mental state that carries on during my experience. On the contrary, I keep passing from a state of perception to one of non-perception, and the reverse. And this is because, when I really see the ocean, what is presented to me is the fact itself—so to say, the world is making itself manifest to me. But when I receive the stimuli on my cortex, what appears to me is not the world itself, but an appearance that happens to match the world. And though it matches the fact, in some way it falls short from the fact (in the same way as a filmed sequence perfectly matches the facts that it captures, but in some way falls short of them). “The disjunctivist conceives of a veridical perception as a relational state of affairs, involving both what is perceived and the perceiver”\(^3\) it is clear that, when I really see the ocean, this relation obtains, otherwise it does not. One may object that this case shows that the disjunctive account is absolutely counterintuitive, as what is in my mind during the whole experiment is a phenomenologically continuous stream, and a theory of perception must be able to account for this. The disjunctivist would answer that it is wrong to think about perception in terms of mental content; perceiving is, on the contrary, entering in a direct relation with the objects in the world.

I want to turn now to an analysis of the example in terms of Searle’s theory. In *Intentionality*, he says that “the visual perception involves at least three elements: the perceiver, the visual experience, and the object (more strictly: the state of affairs) perceived”\(^4\). In particular, “in the case of visual hallucination the perceiver has the same visual experience but no Intentional object is present”\(^5\). Hence, using this theory we should be able to account for my phenomenological stream of experience as having a unique Intentional content—when the experimenter pushes the button, the content of my perception should remain the same. But let us recall the conditions of satisfaction for the perception of the ocean:
I have a visual experience (that there is the ocean in front of me, and that there is the ocean in front of me is causing this visual experience).

We see that these conditions are met just when I properly perceive the ocean, and not when its image is projected into my mind. The reason is that, in this latter case, the second part of the conditions of satisfaction, in the sense of thing required, does not obtain. Note that this is a direct consequence of inserting the existence of the object inside the conditions of satisfaction of the visual experience. The conditions of satisfaction specified in this way are not those of the visual experience, intended as the “highest common factor” between perception and hallucination, but those of successful perception only. In this way, the actual state of affairs in the world comes to determine the Intentional content of our experience.6

At this point, one could argue that even if the conditions of satisfaction of the visual experience, in the sense of things required, do not obtain, that does not mean that we must change the conditions of satisfaction, in the sense of requirements. And, the argument may go, it is just this difference that marks the distinction between perception and hallucination: in the former case the conditions of satisfaction are met, and hence the visual experience is successful, and veridical; in the latter case the conditions of satisfaction are not met, and hence we cannot say that we have a veridical experience. In other words, the Intentional content is exactly the same, what changes is just whether, in fact, certain conditions obtain or not. But I think this solution has a flaw, and will try to point it out.

One of the main advantages of introducing self-referentiality of the conditions of satisfaction, says Searle, is that it manages to account for Twin Earth cases.7 Let us consider the case of two identical twins on Earth and Twin Earth, having type-identical perceptions in type-identical external conditions (light, surroundings, etc.). What is it that makes the two perceptions different, though they are phenomenologically identical? The answer is that, e.g. in the case that the object perceived is a station wagon, “twin number one requires a station wagon causing his visual experience and twin number two requires a station wagon causing his numerically different visual experience”8. The conclusion is: “Same phenomenology; different contents and therefore different conditions of satisfaction.”9 What I want to do is apply the Twin Earth argument to the example with which we are dealing. Of course, we have no problems for the moments in which I have a real perception of the ocean. Given that the above account is correct, my perception is different from the one of my twin on Twin Earth as they are
caused respectively by Earth and Twin Earth-ocean. But what happens in the moments in which I am having a hallucination? Two solutions are possible: either we are having exactly the same hallucination, with the same Intentional content in our minds, or we have two different hallucinations, and the Intentional content must be different. It may be argued that in the case of a hallucination my twin and I are perfectly allowed to have the same Intentional content. If an inhabitant of Earth and one of Twin Earth associate the same sense to the word “water”, but they are referring to two different substances, then a problem arises; but it seems that there is no such issue for an expression like “the king of France”, which has no referent. But I think this account does not work. For, even though I do not properly perceive, I have before my mind an image of the ocean of Earth, while my twin has before his mind an image of the ocean of Twin Earth. My image is in a way dependent on the appearance of Earth-ocean, but it is only a mere case, independent from any kind of nexus, that it matches the appearance of Twin Earth-ocean. We say that the two perceptions of the station wagon are different because they were perceptions of two numerically different station wagons. At the same time, we have to distinguish the two visual experiences of the ocean, as they are visual experiences of two different oceans—even though those visual experiences are, in a way, unsuccessful, i.e., we do not want to say that they are perceptions. Hence we cannot allow that my twin and I have visual experiences with the same Intentional content. If we consider, however, our previous account of conditions of satisfaction of the visual experience, we see that it does not manage to capture the difference. For the relation of causation that those conditions specify does not obtain, and hence it seems that there is no fact in the world that marks the difference between my twin’s and my experiences. In order to obtain a difference in the Intentional content, we must re-state the conditions of satisfaction of the visual experience as follows:

I have a visual experience (that there is the ocean there, and that there is a machine controlled by an experimenter is causing this visual experience).

In this way my twin’s and my visual experiences have different conditions of satisfaction. For my experience is caused by one machine, on Earth, and the twin’s experience is caused by another machine, on Twin Earth. With the previous formulation, we were not specifying the existence of an object causing my visual experience, and hence we were not able to distinguish between the experiences of the twins. Now we have reintroduced such an
object by providing a re-formulation of the conditions of satisfaction of the experience. (The whole argument has been built upon the cases of ‘veridical illusion’, i.e., hallucination which happens to match the world, but I believe that using the notion of indexicality of the Network and the Background, we could manage to extend it to cases of non-veridical illusion. Let us assume that my illusion is related to my Network of beliefs and to my Background, and my Network and my Background are indexically tied to me. Then, though it is type-identical to my twin’s illusion, it is different. I can have the illusion of a ship sailing in the ocean; this illusion will be built on my Network and Background information about ships, and hence will be different from my twin’s illusion of the ship. And this difference should be accounted for when stating the conditions of satisfaction of the visual experience).

One may accuse me that, after all, I am just playing with the notion of causation. He may object that my argument is grounded on the fact that in my example there is a well definite cause to our visual experience, in such a way that the experience is veridical; that cause just substitutes the mechanism of vision. And this is true: what I want to do is just account for the different causal mechanism. But, in order to do this, I need to bring that causal mechanism inside the conditions of satisfaction of the visual experience.

If the above argument is correct, then we must conclude that the Intentional content of perception and hallucination is different, even if my phenomenological stream is identical. Hence, while I am subjected to the experiment, every time that the experimenter pushes the button my state of mind changes, and there is not a unique visual experience carrying over the whole time. One may protest that this result is really incompatible with an Intentional account of perception, yet we can easily come to the same result—a phenomenological stream is broken in different Intentional contents—with a slightly modified version of our example. Let us imagine that the machine of the experiment works differently. It has no camera, and it is directly linked to its correspondent machine on Twin Earth. When the experimenter pushes the button, the machine directly operates on the data present on my brain, switching it with those in the brain of my twin. I still have a unique phenomenological stream of experience, but before my mind there are clearly two different perceptions: first, my perception, and then my twin’s perception. As we have seen, though being type-identical, they have different conditions of satisfaction, which means that the Intentional content in my mind changes every time the button is pushed.

Let us step back to our comparison. We have found that following Searle’s account we do not obtain a unique Intentional content corresponding to my phenomenological stream. On the contrary, the Intentional content of
my experience changes every time the experimenter pushes the button, even though I do not notice any change. And this is just the result we get when we analyze our example in terms of the disjunctive account. Note that I am not claiming that the above argument makes Searle’s theory collapse into a disjunctive theory, as the disjunctivist would not accept that perception is a matter of Intentional content in the mind of the perceiver. Equally, this unexpected outcome does not show that Searle’s account is wrong; I just argued against the claim that we have the same Intentional content in the cases of hallucination and perception. It was not my intention to argue for or against any of the two theories, but just to show that there is a strong similarity in the way they come to explain hallucination cases. This similarity is due to the fact that, rather unexpectedly, Searle’s theory, as does the disjunctive one, turns out to draw a sharp distinction between perception and hallucination. And this, in its turn, is due to the insertion of causation in the conditions of satisfaction of the visual experience. In this way phenomenologically identical visual experiences turn out to be different just because they happen as a consequence of two different facts in the world. Hence, the Intentional content of perception is made strictly dependent on the existence on the state of affairs that causes the perception itself. When the latter changes, the former will change in accordance. A very similar claim is made by a radically externalist account like the disjunctive one: I am in a different state of mind when I perceive and when I am subjected to an illusion; my state of mind is determined by a state of affairs in the world.

NOTES

1 McDowell, *Criteria, defeasibility, and knowledge*, in Dancy, *Perceptual knowledge*.

2 I would like to point out that this example is just a modern version of the ancient argument that a sort of god or demon could create in us the illusion of a certain state of affairs in the world, and that that illusion happened to match exactly what we would normally see. In the present case, the experimenter plays exactly the role of the demon, while his machine replaces its supposed supernatural capacities.

If we had to imagine the conditions of satisfaction of visual experience in a way that they were met by both perception and non-perception cases, they would be simply:

I have a visual experience (that such-and-such is the case).

In other words, the conditions for me to have a visual experience, be it a perception or a hallucination, would be simply that I have such a phenomenological experience, regardless of what causes it. But this, of course, would imply a quite radical break with Searle’s account, as the relation of causation would be no more part of the conditions of satisfaction of the visual experience. I.e., causation would be considered a non-Intentional relation, operating among objects and events in the world.

Notice that these two are not the only possible conditions of satisfaction of my visual experience. Imagine, for example, that the experiment produces a certain damage in my brain, and that, even when the machine is not activated and my eyes are closed, I continue for a few seconds to see clearly the ocean in front of me, in a way that also this experience is undistinguishable from the others, and still continues the uninterrupted phenomenological stream. The conditions for this experience are:

I have a visual experience (that there is the ocean there, and that my brain damage is causing this visual experience).

This example is, of course, rather unrealistic. Nevertheless, I find that it is not unconceivable, and that it deserves serious consideration.
AVIV ROSENBLATT

The Philosopher-King and The King-Philosopher

“But care draws forth the power within,
And cultured minds are strong for good:
Let manners fail, the plague of sin
Taints e’en the course of gentle blood.”

~ Horace, Ode IV.4, Stanza 9
There is a curious omission in the Republic, when after introducing two distinct possibilities for the régime of an ideal city, Plato goes on to discuss only one of them. This crucial juncture takes place at 473d: “There is no end to suffering, Glaucon, for our cities, and none, I suspect, for the human race, unless either philosophers become kings in our cities, or the people who are now called kings and rulers become, in the truest and most complete sense of the word, philosophers – with all those people whose inclination is to pursue one or other exclusively being forcibly prevented from doing so”. The second alternative – the kings becoming philosophers – is never mentioned again in the Republic, despite the fact that it has several important advantages over the Callipolis solution.

However, I think the possibility of a king (or any other political figure) becoming a philosopher and inaugurating an ideal régime is not entirely absent from Plato’s writings. That Plato would seek to provide for such a possibility in his writings should always motivate our reading, since apparently Plato, in real life, tried to assist in realizing such a state of affairs (seventh epistle)¹. In fact, I think there is substantial evidence that suggests that Socrates, in the Platonic dialogues, was himself interested in recruiting political allies. My evidence has to do with characters whose appearance in the dialogues affects the political fate of the philosophical discussion. More importantly, I think it is clear that the Platonic Socrates had one specific person in mind, whom he tried to turn into a king-philosopher: a politician who acquired political power and was then also ‘converted’ to philosophy. In this paper, I will discuss my evidence for the Platonic Socrates’ political interests, and his specific interest in this king-philosopher figure. Finally, I will argue for why the king-philosopher alternative had to be suppressed in the context of the Republic’s conversation.

But first, a word on the king-philosopher. The minimal requirement for a king-philosopher is a politician with sufficient control over the city, who is not a philosopher himself but is receptive to philosophy. A philosopher like Socrates could serve as an adviser to this king, eventually helping him to become a philosopher in his own right². The king-philosopher must have then, before being approached by the philosopher, all the necessary prerequisites of a philosophical nature, but he does not already have to be a philosopher. A description of exactly this situation is given in the Laws 709e6-710d3.

Give me a state under the absolute control of a dictator, and let the dictator be young, with a good memory, quick to learn, courageous, and with a character of natural el-
evation\textsuperscript{3}. And if his other abilities are going to be of any use, his dictatorial soul should also possess that quality which was earlier agreed to be an essential adjunct to all the parts of virtue… I mean the spontaneous instinct that flowers earlier in life in children and animals and in some cases succeeds in imposing a certain restraint in the search for pleasure, but fails in others… This is the innate quality our dictator must have, in addition to the others, if the state is going to get, as quickly and efficiently as possible, a political system that will enable it to live a life of supreme happiness. You see, there is no quicker or better method of establishing a political system than this one, nor could there ever be… – and don’t forget to add ‘lucky’ too, in this one point: he should be the contemporary of a distinguished lawgiver, and be fortunate enough to come into contact with him.

Now what if Socrates were to search among his acquaintances for someone with the potential of becoming such a king-philosopher? Socrates, unfortunately, is a citizen of a democracy, which is only the third most likely to acquire philosophical rule (710e4). He is also devoutly loyal to his native Athens, and so we should expect him to wish to convert Athens, and not some other city – as Plato tried in Syracuse – into the ideal philosophical state. The method, according to the \textit{Laws}, would first of all be to convert the existing régime into a dictatorship, by allying himself first with the most influential elements in the city (τους ἐν τῇ πόλις μεγίστον δύναμεν νομίζειν) (710e9-711a3). In the city of Athens the most likely candidate, it would seem, would be Alcibiades.

The king-philosopher alternative, it should be noted, is also a lot more practical. It could be realized in Socrates’ own life-time, in the person of Alcibiades (or another youth). It would not require the complete revolution which the \textit{Republic} prescribes for Callipolis (541a1), but instead could happen simply through the ascension to power of a politician who subsequently effected a smooth change from democracy to philosophical tyranny. Thus even if this alternative were inferior to that of the philosopher-king, its easier realization could have convinced Socrates to pursue it\textsuperscript{4}. 
On several occasions throughout the dialogues, Socrates firmly opposes having a judge decide between Socrates and the person he is arguing against. In book I of the *Republic* (348a), Socrates poses the problem as an alternative between two styles of argument. One possibility is that Socrates and Thrasymachus will each argue their side, independently of the other’s argument, and then a judge will have to announce the victor. The judge will do this by “counting and measuring the good things mentioned on each side”; clearly, this will be unfair to the philosophical content of the things said. The other alternative allows both sides to be “both jury and advocates at once”, using persuasion on each other, and not resting content until agreement (ομολογία’) is produced. Thus Socrates presents judgment here as something external to philosophy, something that is better avoided.

Socrates’ attitude towards judgment as a means of deciding a philosophical argument is uniform throughout the dialogues. However, on several occasions a judge does intervene in the dialogues, without Socrates’ apparent knowledge or agreement. The role of ‘judge’ can be supplied by any authority – it is an external power that regulates the philosophical discussion by appeal to its authority alone, rather than the principle of ομολογία’. Another version of this role is, then, the philosopher’s ally, mentioned by Socrates in the *Republic* 496d1.

One of the main issues discussed in the *Protagoras* is the limits and politics of a philosophical discussion. Socrates and Protagoras disagree over how long Protagoras’ answer should be, and who has the right to determine this (329b; 334e). Socrates is ready to end the conversation, because an agreement on its rules cannot be reached (335b; again at 348). Socrates’ readiness to leave is an indication that he is unwilling to participate in a lop-sided argument, in which the two sides don’t share a common ground of agreed principles, and don’t consider agreement between them as the only way to proceed.

At this point, Alcibiades “jumps in” (336c), to defend Socrates’ philosophy on Socrates’ own terms. Alcibiades shows himself aware of the responsibilities of a speaker in a Platonic dialogue. In addition, Alcibiades calls the matter for a public vote – he acts like a politician, and specifically a politician who has taken it upon himself to defend philosophy’s sovereignty. Finally, Hippias, who was seated on a θρόνος earlier (315c) passes judgment, suggesting that they choose a moderator, an idea which Socrates rejects straightaway (338b). Ten Stephanus pages later, the conversation is about to dissolve again, before Alcibiades again intercedes to save it. This time
his role is even more partial: he questions whether Protagoras is behaving well or not, after which Protagoras is said to be insulted (348c1).

Alcibiades is thus quite valuable to the conversation: “In terms of the philosophical drama as a whole, however, [Alcibiades’] interventions take place at crucial points when the conversation threatens to disintegrate, and they save it from collapse” (Friedländer, Vol.2, 6). Furthermore, Alcibiades is very bold in his defense of Socrates’ philosophical principles. He proves himself ready to take liberties with an older, much-respected man to make sure the conversation doesn’t fall apart. Socrates acknowledges neither of Alcibiades’ two efforts when they are made. In the introduction to the dialogue, however, Socrates does show that he noticed this.

The spiritedness with which Alcibiades came to his defense might have convinced him that Alcibiades could be politically valuable to philosophy. If the action of the Alcibiades takes place after that of the Protagoras (both take place around the year 433 (Taylor 64; Denyer 189) the description of Socrates in the Protagoras 309a2 as still in pursuit of Alcibiades would suggest that they are not yet lovers), this might have been the occasion that made Socrates decide to enlist Alcibiades as a pupil and try to make him into a king-philosopher. This bit of speculation aside, the important issue is that the Protagoras begins with a mention of Alcibiades, which suggests that he is a key character – just as the Euthydemus’ closing with Isocrates will reveal his importance.

Isocrates

The tale of Socrates’ defeat by the sophist brothers in the Euthydemus is framed by Socrates’ meeting with Crito. After hearing Socrates glibly recount the event, Crito mentions that he had already heard the story, from an unnamed, unnoticed observer, “someone who has a high opinion of himself for wisdom and is one of those clever people who write speeches for the law courts” (304d). Socrates is very interested in identifying this observer (apparently Isocrates) (Complete Works 708), and seems to figure it out following Crito’s description: “Now I understand – it was about this sort of person that I was just going to speak myself” (305c)7.

In any case, Socrates’ curiosity about the identity of the observer fits into the larger question in the dialogues of whether a speaker must express his own views or not. Socrates is often preoccupied with the question of who voiced a view, rather than what the view itself claims. A useful parallel
to the case in the *Euthydemus* can be found in the *Charmides*. Charmides seems to recognize this Socratic preoccupation when he says: “What difference does it make, Socrates, from whom I heard it?” (*Charmides* 162c). In both passages, Socrates’ next words are directed towards the original author of the view, rather than the interlocutor who voiced it. In the *Charmides* this is made evident in 162c, when Socrates notes that Critias’ manifest anger helped confirm his earlier suspicion that Critias was the author of the view. In the *Euthydemus*, too, Socrates’ response immediately centers on disqualifying Isocrates, the author of the view. The focus of the dialogue suddenly shifts at this point to the person of Isocrates, whom until now we had not even known was present. Furthermore, Socrates’ interest in the author of the view expressed by Crito at the end of the *Euthydemus*, the fact that he proceeds to an *ad hominem* attack on this unnamed person, and the strange remark he makes that he was “just going to speak” of this sort of man – these are all signs that Socrates has been deeply concerned with Isocrates himself for a while.

The sudden concern with Isocrates here suggests to me that a reinterpretation of the dialogue as a whole, and particularly Socrates’ political failure in it, is in order, and should be based on this new information. We are now told that Isocrates was present all along at the discussion. Isocrates seems to play here the role of a judge, passing judgment at the end of the dialogue on all that transpired in it (304d-305b). Having seen Socrates get lost in the brothers’ cul-de-sac sophistry, Isocrates now criticizes philosophy in general as being of no value, as just “chattering and making a worthless fuss about matters of no consequence”. And finally, he considers this activity ignoble, so that Crito would have been embarrassed (305a2) for his sake. Isocrates remarks that Socrates should not degrade his philosophy by talking with such people, who “care nothing about what they say, but just snatch at every word” (305a). Since Socrates appears so interested in the question of who made these comments, we should try to see what they reveal about their author, Isocrates. Without actually participating in the discussion, Isocrates shows that he cares about the fate of philosophy. He does this by revealing his presupposition that philosophy *should* be a noble thing. Socrates, in his subsequent exposition of Isocrates’ views, identifies Isocrates as one who considers himself a philosopher of some sort (305d).

I would argue that the best explanation for why Socrates deliberately lost the argument with the sophists is in order to elicit such a reaction from Isocrates, whom he must have known was present in the audience. By failing so miserably, and failing to acknowledge his failure, Socrates knew he would rouse Isocrates’ indignation over the humiliation of philosophy.
Socrates uses, then, a strategy quite opposite to the one he used in the *Protagoras*. This difference in strategy reflects the different characters of the two judges, Alcibiades and Isocrates. Alcibiades, who was only about seventeen at the dramatic date of the *Protagoras*, did not have as lofty a conception of philosophy as Isocrates did. Callias describes him there as always wanting “to be on the winning side of a good fight” (336e). To be inspired by philosophy, Alcibiades needs to see it victorious; Socrates cannot allow himself to take a beating from a sophist. I think Socrates’ entire style, how he carries himself in this dialogue, can be traced to Alcibiades’ presence: Socrates presents himself as a busy, important man, eager to quit the conversation (362a). Having torn Protagoras to shreds (it is Protagoras who backs away from the conversation, with a parting compliment to Socrates, at 361e), Socrates makes a triumphant exit; this is the absolute contrast to his humility at the end of the *Euthydemus*. Certainly, both Isocrates and Alcibiades are being tested by Socrates to see if they will stand up for philosophy. Both men present political solutions to situations in which Socratic philosophy fails – Alcibiades puts to shame and pressures reluctant interlocutors, while Isocrates shows that he has his own conception of whom a philosopher should talk to and how he should carry himself about.

*Alcibiades*

There are many ways of misunderstanding the relationship between Alcibiades and Socrates. First of all, the relation between Alcibiades and Socrates was not erotic, in the common sense of the term. If we are to believe Alcibiades’ confession at the end of the Symposium (and I see no reason not to), Alcibiades was quite shocked that Socrates would reject his sexual advances, and seems to have not yet recovered from that shock. Socrates’ abstinence in their relationship made him more mysterious, and more admirable, in Alcibiades’ opinion. Furthermore, even if it were argued that Socrates was ‘into’ non-sexual erotic relationships with young boys, there still remains the question of why he would choose Alcibiades. There were plenty other handsome boys available, whom we know Socrates was attracted to (*Charm*. 154b-c). Socrates’ fascination with Alcibiades was one of a kind (*Alc*. 103a-104d); and since Socrates was a philosopher, we don’t want to explain this as just a random infatuation. Obviously, somebody who accepted the general protreptic theory might say, Socrates was interested in turning his young interlocutor into a philosopher. But this still does not explain why Socrates chose Alcibiades and not, say, Theaetetus. *Theaetetus*
looks like better philosopher-material; he is gentle, willing and quick to learn (Theat. 144a), and doesn’t claim to know more than he does (e.g., 148e). He seems, in fact, to resemble Socrates. Alcibiades, on the other hand, is proud, confident in his abilities and knowledge (103b-104a), and doesn’t look like he could make a serious student. And of course, in hindsight, Theaetetus apparently did become a mathematician (Complete Works 157), which is perhaps closer to a philosopher than Alcibiades ever reached.

At this point, the believer in the protreptic theory is likely to endorse what may be called the standard-redemptive view: Socrates was always more interested in arguing with people who felt confident in their false knowledge, who were farthest from having adequate philosophical qualifications. This view can be supported by the gadfly analogy which Socrates makes in the Apology. However, I don’t see any reason why Socrates should take up this ‘redemptive’ approach and deliberately seek out those who are farthest from philosophy, or the greatest challenge to convince. Socrates became Alcibiades’ lover, and so expended an extended amount of time on him, which he could have used better on several other young men who were more likely to end up as philosophers. Thus even if the redemptive view claimed that someone like Theaetetus didn’t require much help from Socrates (which is likely true), while Alcibiades did, I still think it cannot give a sufficient explanation for why so many others are passed up.

The passage in the Republic about the natural obstacles preventing a talented youth from capitalizing on his philosophical potential (490e-6a) bears directly on the question of why Socrates would approach an Alcibiades. It seems, in fact, that this passage alludes directly to Alcibiades, when it speculates on what might befall a youth of strong character, “especially if he chances to be from a big city, is rich and noble in it, and is further, good-looking and tall? Won’t he be overflowing with unbounded hope, believing he will be competent to mind the business of both Greeks and barbarians, and won’t he, as a result, exalt himself to the heights, mindlessly full of pretension and empty conceit?” (494c-d). Such a youth couldn’t possibly succeed in becoming a philosopher.

This passage (Rep. 490e-6a), therefore, only thickens the mystery of why Socrates would approach Alcibiades, who is so likely to become corrupted. Hatzfeld also notes that it is strange that Socrates would not have anticipated his Alcibiades’ too-powerful attraction to politics. He then adds: “mais jamais, ni chez Platon – et pourtant l’occasion eût été bonne dans le Banquet –, ni chez Eschine, ni chez Xénophon lui-même, le maître [i.e. Socrates] ne proteste contre cet espèce d’abus de confiance qui aurait consisté à vouloir utiliser son enseignement pour des fins diamétralement opposées
à celles que poursuivait celui-là même qui le dispensait” (38-9).

One could argue that the action of the Republic happens after the action of the Alcibiades, and so Socrates’ words here are a sort of regretful accounting, taken from bitter experience. But if Socrates is speaking from experience and thinking just about Alcibiades here, his generalization about ‘such a person’ loses force. And still, it remains unclear why Socrates had to wait for such experience – why he didn’t realize all this a-priori. It seems to me, then, that the passage from the Republic actually proves that Socrates did not approach Alcibiades for the purpose of making him into such a philosopher as is described in this passage. If megalomania and dreams of international politics are traits that characterize the most extreme corruption of a gifted young nobleman, Socrates would not have himself suggested these ideas to Alcibiades on their first meeting. By comparing the young Alcibiades to the prince of Persia, Socrates was deliberately putting this idea into Alcibiades’ mind. Of course, Socrates would not suggest these dangerous ideas to Alcibiades for the purpose of later dispelling them; this would be pointless and too risky – after all, “it is from people like this that those who do the greatest harm to cities and individuals come…” (495b3-4).

Further, the apparently extraneous description (Alc. 121e5) of the four specialist teachers of the Persian prince seems to prefigure Socrates’ own roles in regards to Alcibiades. Socrates, in the end, supplied Alcibiades with the training that each of these teachers should give: As teacher, Socrates obviously instructs Alcibiades about wisdom and justice. But as lover and brother in arms, Socrates also provided Alcibiades with a shining example of courage and self-control (Symp. 219d5). The implicit message is clear: Socrates first suggested to Alcibiades in this dialogue what role he could play in international politics, and proceeded to train him in every way required for realizing this goal.

Thus the passage in the Republic (491e-496a) seems to rule out the possibility that Socrates would approach Alcibiades for protreptic purposes. The dangers described in this passage render Alcibiades a poor candidate for a philosopher; and Socrates, in the Alcibiades, seems to have only quickened these dangers. The only explanation can be that Socratic wanted to make Alcibiades into his pupil and beloved for other purposes. I shall now offer a better interpretation of the relationship between Socrates and Alcibiades which indicates what Socrates was trying to make of Alcibiades.
One of the remarkable points which Alcibiades makes in the Symposium, looking back on his relationship with Socrates, is that he discovered to his surprise that their erotic roles had been reversed: he had become Socrates’ lover\(^\text{14}\). The erotic relationship is related metaphorically to the relation between the statesman and his city. In the Alcibiades, Socrates expresses his fear that Alcibiades will turn out a δημεραστζ (132a). Denyer’s commentary on this line refers to the Gorgias 481d, where Socrates accuses Callicles of being just such a lover of the people. Denyer glosses the term as a spineless politician who “will adapt himself to all the fickle changes of public opinion as the lover does to all the moods of his beloved” (226). Socrates, in any case, wishes that Alcibiades not become this kind of a lover of the city. I would suggest that the reversal of erotic roles between Socrates and Alcibiades is supposed to mirror the reversal of roles that should happen between Alcibiades and Athens. As an aspiring politician, Alcibiades needs to court the city’s favor. But Socrates wants Alcibiades to go beyond that stage, and actually become, in some sense, the beloved of the city. This presumably means that the people will now seek to find favor with him\(^\text{15}\).

The move from lover of the city to beloved is therefore significant for the politician in two ways: First, the politician no longer has to be a δημεραστζ. He does not have to alter his views to match the crowd, and in fact can now retain a philosophical composure: his philosophical principles will no longer be corrupted by his political involvement. And second, now the people will become his lovers, which means that they will be more ready to alter themselves based on his requirements or whims. The politician who is philosophically motivated will have greater opportunity, in such a case, to influentially apply his philosophical principles to the people.

We see that the move from δημεραστζ to beloved of the city appears perfectly suited for a philosophically-minded politician. It allows him to preserve his philosophy from the negative effects of political action – from precisely those dangers described so vividly in the Republic (491-6). Furthermore, if he succeeds in effecting this reversal, he will be able to acquire an unprecedented level of influence over the people with his philosophy. I think Socrates meant for Alcibiades to accrue these advantages by becoming the beloved of the people. After all, no other politician can be said to have been as beloved by his city as Alcibiades was when, having betrayed them, they still yearned for his return. As Aristophanes summed up: ‘they long for him, they loathe him, they want to have him’ (Frogs 1425, quoted from Denyer 3).
Socrates, as determined above, did not need to convert Alcibiades into a philosopher right at the start of their relationship. Rather, it was sufficient that he should learn how to acquire political power, at the same time becoming loyal to Socrates, and that he be concerned with improving himself and becoming more philosophical (I take this to be the result of την δημωδή σοφροσύνην mentioned in the *Laws* 710a5-6). Alcibiades’ loyalty to Socrates was secured by the role reversal, which made Alcibiades fall in love with Socrates. Next, there are several ways in which Socrates tried to teach Alcibiades the importance of philosophy for his political future.

The first, most obvious, is just by experience and imitation: Alcibiades would see Socrates charming him into changing roles. Alcibiades would know how it feels to be a lover, and also would witness Socrates’ crafty reversal of roles and perhaps learn how to imitate it. Secondly, Alcibiades could become convinced that philosophy has this power to turn beloveds into lovers. He might realize how useful this could be, and how useful Socrates could be for this move. Thirdly, and most importantly, the move from δημοκράτις to beloved of the city depends on an ability to manipulate the beloved. I will give my interpretation of one passage from the *Alcibiades*, in which we witness Socrates giving Alcibiades justification for such a manipulative move.

Alcibiades submits that he considers the knowledge of good advice (ευβουλία) to be the knowledge one needs in order to rule in a democracy (*Alc.* 125e6). Socrates and Alcibiades at first agree that friendship and agreement are the marks of a successful city (126c). However, the example of the relation between the sexes shows that a harmonious situation does not require agreement between the two sides, but rather each doing his or her own job, since “some subjects are women’s subjects and some are men’s subjects” (127a). Likewise, Alcibiades confesses that he believes a city is well-governed when different groups each do their own work. But Alcibiades is confused about whether the principle of ‘one man, one job’ means there is friendship and agreement in the city. At this point (127d), Alcibiades becomes frustrated, and Socrates allows the conversation to shift to the question of Alcibiades’ education.

The radical conclusion, which Socrates did not state directly but was obviously using to prod Alcibiades’ replies, seems to be that justice does not require agreement among the different groups, and that friendship is not necessarily the same as agreement. Socrates’ leading questions reveal that he’s trying to get Alcibiades to consider these ideas: “Can there be friendship...
without agreement? Can there be any agreement when some know about the matter and others don’t?” (127c2-3). Socrates seems to be pointing Alcibiades towards a few of the anti-democratic principles of Callipolis – especially the principle that the need for agreement is overcome by other, more important considerations of justice in the city.

Finally, in his last word on the subject, Socrates reminds Alcibiades of the issue which opened this discussion: ευβουλί'α (127d1-2). But by tying in the issue of the statesman’s role, Socrates is suggesting a reinterpretation of the previous discussion. Socrates is suggesting that ευβουλί'α can be used for purposes other than fostering agreement and friendship. Or rather, since justice and the best end for the city do not require agreement and friendship, the statesman practicing ευβουλί'α can feel free to shed his pretences to democratic unity when it suits him.

Until now, Socrates seemed to be talking on a completely different level than Alcibiades – Socrates was describing what knowledge could actually make the city good, while Alcibiades was only interested (124e-125e) in the politician’s limited role in the city, which mainly consists in acquiring and maintaining the public’s support. Socrates seemed to be disregarding the possible rhetorical/manipulative uses of ‘good counsel’ – that the agreement it fosters can be based on a lie. While superficially ignoring Alcibiades’ concern with his role as politician and speaking of what makes a city just, Socrates was actually, covertly, pointing out to Alcibiades a justification for anti-democratic manipulation.

The Denial of the King-Philosopher

My account of the Republic may be challenged to explain why Plato or Socrates would go to such lengths to describe a city with such a fundamental philosophical flaw (the typological reduction and its consequences for education), and why Plato would deliberately suppress the possibility of a king-philosopher. My answer to the first question does not have to be too belabored, since the Republic is not rendered irrelevant if Callipolis is not the ideal city for philosophers. In any case, the character of the interlocutors in the Republic supplies the answer to the first question, and by extension, to the second as well. Reeve poses the question succinctly: why are Glaucon and Adeimantus needed to resume Thrasymachus’ argument for injustice (Reeve 3; 33-41)? Their unique character17 shape the discussion of Callipolis in many ways. I will here discuss only what I consider their most important effect on the conversation, and how it rules out the possibility of
introducing a king-philosopher.

It has been remarked by commentators that Glaucon and Adeimantus’ character quite closely corresponds to that of the warrior class. Their support for the principles of Callipolis which Socrates lays down, then, amounts to a willingness to fight for these principles – as the warriors in Callipolis should. Glaucon and Adeimantus are to fight on two fronts: in defense of philosophy in general against the charge that it is politically useless, and by showing a willingness to make ideal sacrifices for the sake of the ideal justice of Callipolis. This first battle is announced by the repeated references to the common view, that those who indulge in philosophy are useless to their city. The ultimate response to this view is that philosophers would be most useful to the city, if they were allowed to rule. The second battle takes place as Glaucon and Adeimantus agree to conditions that they find counter-intuitive or unpleasant in Callipolis – for example, absolute communism. As he gleans their affirmation of such principles, Socrates proves to the brothers that philosophy can be valuable for political solutions, and inspires them to defend it. Now just as Glaucon and Adeimantus correspond to the warriors of Callipolis, Socrates corresponds (certainly in Glaucon and Adeimantus’ view) to the philosopher-king. Their assent to the principles of Callipolis is translated into a loyalty to Socrates: a class loyalty, of the warriors to the philosopher-kings.

Now in real life, Socrates is not leaving the brothers any immediate path towards political action. If they are really interested in a just city, they cannot act alone; they need an entire class of philosophers on their side. If they now insist on having a political coup on their own, they will not be able to justify it philosophically. A basic premise of the discussion in the Republic is that only when the wisest people (the philosophers) are in control can justice come about. The important point then is that Glaucon and Adeimantus have no direct way of starting to bring about Callipolis; the question is not whether Callipolis is possible under different circumstances. Burnyeat’s “Utopia and Fantasy” (in Gail Fine: Plato 2), therefore, is misdirected. Burnyeat insists that the ideal city is possible metaphysically if not historically. If the ideal city is not a practical possibility, Burnyeat writes, then any discussion of it is just “idle day-dreaming”. At the end of his article he writes: “[The practicability] claim is made, therefore, not with a view to what could be done in Athens or Syracuse, tomorrow or the day after, but from a standpoint of both temporal and geographical neutrality. It is a claim about human nature as such, not about the men and women we know... The whole Republic is an exercise in the art of persuasion, designed to lead us from here to there. The ideal city is built in our
imagination by persuasive argument, in such a way that successful persuasion in the world of imagination guarantees the possibility of success in the actual world” (308). By focusing on its metaphysical practicability Burnyeat ignores the immediate political implications. For Glaucon and Adeimantus, the discussion in the Republic does not provide any visible political plan, and in fact removes any philosophical justification which they might have been using to support their own political agendas.  

There is a second distinguishing trait of Glaucon and Adeimantus, which Socrates would like to purge them of. Besides being politically ambitious, the brothers are also philosophically pretentious. Glaucon and Adeimantus belong to a type of interlocutor that is averse to remaining in aporia. They are action-oriented, eager for conclusions and positive doctrines. Most importantly, they are confident enough in their philosophical ability, to be unwilling to let anything remain beyond their reach; they never consider that they may be inadequate to the task. They demand the whole story from Socrates – a convincing account of perfect justice. They want to ‘accomplish something’ during the night’s conversation (497e1; 506d2). Socrates tries to rid them of their pretentiousness through the use of fantastic imagery.  

In various parts of the Republic devoted to the structure of Callipolis, Socrates describes picturesque details, the philosophical relevance of which is sometimes questionable. The suggestion that children be taken on field-trips to witness battle, for example, doesn’t seem crucial to the characterization of Callipolis, while so much information that is much more important is missing. Another questionable item is Socrates’ argument for the emancipation of women. Friedländer describes the details of this program as “amusing and even grotesque” (Vol.3, 103). 455d8-e1: “Natural attributes are evenly distributed between the two sexes, and a woman is naturally equipped to play her part in all activities, just as a man is – though in all of them woman is weaker than man”. It does seem that it would fit the ‘one man, one job’ line of reasoning to allow women to take any job. But this argument evades an important question. It proves that women have the same nature as men, only weaker; but Socrates never proves that it is therefore better for Callipolis to emancipate women. Socrates establishes that there is no essential difference between men and women, but this does not mean that there is not a difference of degree between them (454e), which should be considered as well. It also could not be said that women should be liberated just because everybody in the city needs to belong to a category that preserves the city-soul correspondence; since, after all, there will still apparently be slaves in the city (469c; cf. also Vlastos, “Does Slavery Exist in Plato’s Republic?”).
If slaves could remain in the city, engaged in their traditional roles, without disrupting the city-soul correspondence that makes Callipolis, then we are lacking an argument for why women must be emancipated.

In any event Socrates does not present the emancipation of women as necessary for the success or justice of Callipolis. The same can also be said about the role of women in war (and the problems this poses for child-production, which are never addressed in the Republic) and about most aspects of the radical communism. These scenes, however, also happen to be the parts of the story of Callipolis that are most shocking for the interlocutors. At the beginning of book V, Adeimantus and Polemarchus describe communism as the most controversial part of Socrates’ theory. Socrates seems to enjoy the provocation, as when he asks Glaucon to picture to himself the naked old women at the gymnasium (“Τι...γελοιοτατον αυτων οραιζ;” 452a10).

These scenes which Socrates conjures up for the interlocutors are not just ridiculous, or strange – they are fantastic. They make Callipolis not just a perfect city, but also an exotic place. But this magic, while dazzling the interlocutors for the duration of the conversation, is not necessarily appealing to them. I’m quite sure that Glaucon, in his next nightmare, will be wrestling with naked old women at the palaestra. It may flatter Glaucon and Adeimantus’ pride to agree to make such great sacrifices for the sake of the city, but they have probably also become quite apprehensive of the further wild and seemingly arbitrary innovations which Callipolis might bring. The beautiful city is too fantastic. The effect of this on Glaucon and Adeimantus is to make them fear the consequences of unbridled philosophical discourse. Socrates’ use of fantastic imagery, then, seems designed to induce philosophical vertigo in Glaucon and Adeimantus and thus purge their philosophical pretensions.

The characterization of Glaucon and Adeimantus clears the ground for the typological reduction that plays such an important part in the Republic. The typological reduction comes about as an attempt to ward off the risk of people like Glaucon trying to have it all. Spirited people like Glaucon and Adeimantus must be set within limits and not allowed philosophical justifications for their political schemes. They must put all their trust in the philosophers running the city, and not desire rule themselves.

The typological reduction is not only at work in the Republic within Callipolis; it also governs what Socrates can and cannot say to his interlocutors. Socrates cannot mention the possibility of a king-philosopher, the man who combines the rational capacities of the philosopher with the spiritedness of the politician, and thus defies the typological reduction. There are two
points at which the omission of this possibility must be deliberate. Socrates gets closest to announcing this possibility when he dramatically pronounces that there will be no justice until kings begin to philosophize or philosophers come to rule (473c-e). The Republic discusses only the second alternative: how to get philosophers into office. But instead of exploring, or even refuting, the possibility of the king-philosopher, Socrates prefers to pass it up. Similarly, in the discussion of the tyrant in book IX, Socrates does not allow for the possibility of a just, philosophical tyrant. Yet we know that Plato did consider such possibilities. The second interesting omission concerns the person of Dion. Friedländer notes, following Adam’s commentary on the passage, that “Dion rightly is identified as the man of noble and well-bred nature who was saved by exile (496b)” (Vol.3, 111). The startling fact is that Plato here, in a passage meant to prove the impossibility of the coincidence of politics and philosophy, mentions in passing a person who was philosophically trained and still had political aspirations. Plato thus manages to sneak in a refutation of the message which the rest of the passage succeeds in delivering to Glaucon and Adeimantus.

The king-philosopher must be made to appear impossible, so that Glaucon does not try to become one. That’s because Glaucon and Adeimantus, as argued above, are ambitious politically and arrogant philosophically – they scheme for power and they consider themselves capable of any philosophical understanding. Socrates wants to dissuade them from idealist politics, and so must deliberately suppress the possibility of a king-philosopher. Socrates seems then to do just what he shouldn’t – instead of denying the possibility of a union of philosophy and politics in one person, Socrates grants it, in the person of the philosopher-king ruling Callipolis. The philosophers of Callipolis, who have been brought up from the warrior class, seem to present just this union. In fact, however, Socrates just uses this to bait Glaucon and Adeimantus, since he limits this possibility to the conditions of the Callipolitan system. Socrates had to affirm this possibility in some way, in order to secure the brothers’ loyalty to philosophy, since the brothers now see a way in which philosophy can be useful. However, he had to remove this possibility from daily life, and fix it only in the theoretical Callipolis, in order to curb Glaucon’s ambition.

What accounts for the difference in Socrates’ attitude towards Alcibiades and Glaucon? Alcibiades’ attitude towards Socrates swings between resistance and suspicion on the one hand (in the beginning of the dialogue bearing his name), and full submissiveness on the other, culminating in self-renunciation and ‘conversion’ (127d). Their relationship was intensely personal, as also reported in the Symposium. Alcibiades’ passion – as well
as the extent to which he could commit himself personally – marked him off for Socrates as a potential philosopher and ally. The discussion in the *Republic* does not reach that kind of intimacy. More importantly, the theoretical discussion of justice is never applied to the lives of the interlocutors in the *Republic*. This is perhaps inevitable, since the discussion takes place amongst a larger company. But the difference in tone still marks the different purposes to which Socrates could put the different interlocutors.

My conclusion, therefore, is that the *Republic*’s governing principle, that of the typological reduction, derives from the character requirements of the interlocutors in the dialogue. The reduction then in turn necessitates the ideal construction of a city which is not the best for philosophy or philosophers, yet still manages to fulfill several important criteria of justice. A certain circularity in the *Republic* attests to its organic unity of matter and manner: Glaucon and Adeimantus’ characters impose the typological reduction on Callipolis and on the conversation; yet this typological reduction is what gives legitimacy, or clears the ground for the confining of men like Glaucon and Adeimantus into a single group. The *Republic* on the whole is interested in exploring different notions of justice, and for this purpose the ideal construction of Callipolis is useful. But Callipolis should definitely not be taken as the ideal philosophical city.

The question may still be raised why the possibility of a king-philosopher is not introduced in other dialogues which don’t star Glaucon and Adeimantus. I would answer that the same reasons that Plato had not to suggest the king-philosopher in the *Republic* hold for every other dialogue. The danger posed by characters such as Glaucon and Adeimantus is always present; Plato would rather have his readers search for a philosopher to crown as king, rather than declare themselves kings, hoping that enlightenment will follow. The education of the king-philosopher is a sensitive subject; the educating philosopher must tread lightly, to fan his pupil’s ambition while not letting it get out of hand, thus losing him as a philosopher. It was a delicate task that could only be performed by a philosopher personally, but was too dangerous to commit to writing, according to the reasoning and the experiences described in the seventh epistle. Alcibiades is the only exception, the only time we get to see Socrates really offering his service to a potential king-philosopher. But Socrates was willing to invest the effort in Alcibiades, and would probably not do the same for any other Athenian he knew.

If I’m right about the reasons for the suppression of the possibility of the king-philosopher, we can learn from this an important point about cross-referencing in the Platonic dialogues. Plato in the *Republic* shows...
why he must deny the king-philosopher: because of people like Glaucon. But this denial is meant to hold everywhere else too. Plato does not want to have to develop characters like Glaucon and Adeimantus for every dialogue; he does this once and for all in the Republic, and then expects the reader to know that this same principle applies to all other dialogues. This is an important argument in favor of the thesis that all dialogues should be read in cross-reference; an assumption or conclusion in one should be kept in mind in reading another.

Another good reason to play down the possibility of the king-philosopher is simply to protect his own good reputation, and that of Socrates, which was already tarnished by his association with Alcibiades. In a world that knew well the dangers of tyrants – philosophical or not – Plato preferred to be seen as pushing for a possibility that was seen as less of a threat: the totalitarian state of Callipolis. Plato couldn’t envisage a different world – our world – in which this kind of a state is even more feared. Because he tried to pose as a moderate in the ancient world, Plato earned the title of the ‘father of totalitarianism’ in the modern world.

NOTES

1 Which is why it is simply astonishing that, to the best of my knowledge, no writer on Plato has sought to discover Socrates in the Platonic dialogues engaged in any such positive political enterprise.

2 Without becoming a philosopher in his own right (i.e., by only being a supporter of philosophers), the king would be vulnerable to other, anti-philosophical influences, such as those described in the Republic 494d-e.

3 The qualities described in this passage are almost exactly those ascribed to the philosophers of Callipolis in the Republic 487a, as Strauss notes (City and Man, 93).

4 Mario Veggetti has also pointed out the practical advantages of the king-philosopher option (145).

5 The capacity of judging (δικαστικη) is given a decidedly political cast
when it is related, along with generalship and a branch of rhetoric, to the expert knowledge of statesmanship in the *Statesman* 303e7-304a2. Socratic dialectic is defined by this progression by argument and agreement. On this subject, see Mittelstrass, “On Socratic Dialogue”, esp. pp. 129-132.

6 I’m leaving aside the question of whether Socrates is unfair in insisting on *his* rules. Even if he is, his insistence on the other party’s agreement with these rules is more crucial.

7 I think that Crito’s description at 305c is specific enough that Socrates could figure out who this person was, and not only of what ‘sort’ he was. Crito does not actually name him, nor does Socrates, out of a kind of etiquette, I believe – Crito’s respect for something that was said to him privately. It seems to me, therefore, that Socrates does recognize that this person is Isocrates; but I admit that this is debatable.

8 Plato manages here to intertwine the dramatic exchange with the subject matter: the view being discussed is that temperance is minding one’s own business. And this is precisely what Critias either cannot do – when he finally interrupts the conversation – or rather should have been doing all along – he should have stood up from the start for his view.

9 Socrates’ preoccupation with the author of any view is related to his customary insistence that an interlocutor only argue for a view he holds. Vlastos in his essay “The Socratic Elenchus: Method is All” (In Socratic Studies, p. 8-9) offers three motivations for this. By forcing interlocutors to stick to their own views, Socrates (1) forces them to be more honest in their arguing style, (2) demands seriousness and commitment to the views expressed, and (3) is better able to convince the interlocutors to change their lives. Developing the question further, we could also ask how having the interlocutor speak for himself helps *Socrates* prepare a different argument based on who he is talking to. This perspective could then also explain Socrates’ preoccupation with finding out the author of any view – he does this not only to demand seriousness or honesty on the interlocutor’s part, but more importantly in order to respond directly to the author of the view.

10 But even if I am not granted that his failure was deliberate, I think my point still stands. The point is that Socrates shows, at the end of the *Euthydemus*, that he is very much concerned with Isocrates’ concern for philosophy.

11 My interpretation cannot discern whether Socratic strategies are used to examine an interlocutor’s character (i.e. that Socrates does not know it already) or to cause an effect based on prior knowledge of that character.

12 One who thought the *Alcibiades* spurious would not have this option, and might take the opposite approach: the *Alcibiades* would be merely an
illustration of the principle found in the *Republic* 490-5. But this explanation, again, entails that the author of the *Alcibiades* failed to account for why Socrates did not realize that there was never hope for Alcibiades. And this failure entails the same problem as that which arose for the view that Socrates is speaking from bitter experience: the problem of how Socrates can present a view with general force based on one example.

13 105b4-7: “Having shown that, you’ll be the most influential man in the city, and if you’re the greatest here, you’ll be the greatest in the rest of Greece, and not only in Greece, but also among the foreigners who live on the same continent as we do”. Also 119d6-d: “wouldn’t you keep your eye on your real opponents and not on your comrades, as you’re doing now”.

14 It is interesting to note that there seems to be a hint that another relationship is reversed – that of the intellectual midwife. Socrates in the *Theaetetus* describes his role as midwife (149a-151b) as revealing the inner contents of the interlocutor, bringing them out to light for examination. This is just what Alcibiades appears to have done at the end of the *Symposium* – he discovers Socrates is like those statues, full of jewels. This would suggest also that Alcibiades is taught how to philosophize by the reversal. See also “Socrates the Beautiful: Role Reversal and Midwifery in Plato’s Symposium”.

15 Again, there is a relation here to the *Laws* 711b4-c8, where Socrates describes how the people will emulate their leader.

16 The irrelevance of the agreement of the governed party is discussed in the *Statesman* 292e-3a, 295e-297c.

17 “Apart from Glaukon’s greater promise, however, the brothers are marked less as intrinsically different than as variations on the same character type, with Glaukon serving as a superior and more fully developed example of that type” Blondell 220.

18 Strauss, *City and Man*, 112: “Glauc, the most spirited speaker in the work, who as Spiritedness incarnate comes to the assistance of Reason in the founding of the just city”. Other commentators classify them as potential guardians of the city: Blondell, for example, includes a thorough discussion of their potential (203-219). One reason for this is that “In virtue of their role as legislators, they are equated with the ruling class of philosophers, thus eliding the gap between politics and its representation (497cd)” (212). Blondell and Strauss disagree on this point, therefore, whether as founders, Glauc and Adeimantus are more like guardians or auxiliary warriors. It seems like Strauss is more interested in enforcing a strict analogy – if Socrates is the philosopher, the brothers must be the warriors – viewing Socrates as the teacher and the brothers as pupils. Blondell, on the other hand, considers the dialogue between them as a shared enterprise, and so gives them equal
credit as founders. Blondell admits, in any case, that they are still threatened by a tyrannical nature (212-3), which would seem to correspond more to the auxiliaries than the true philosophers.

19 Glaucon explicitly describes himself as Socrates’ defender at 474d. Denyer (94, 139) quotes Xenophon’s *Memorabilia* 3.6.1 and 3.6.15 on Socrates dissuading Glaucon from speaking at the assembly while too young.

20 Eric Brown seems to hold the same view: “But in and of itself the huge gap between Kallipolis and ordinary politics defines no fixed Platonic political program” (Brown 14).

21 Again, my interpretation of the use of fantastic imagery in the *Republic* here differs from Burnyeat’s approach. Burnyeat concludes, in his “Utopia and Fantasy”, that the *Republic* uses comedy and fantasy in order to break down our “partial, parochial perspectives of everyday existence” (308). I agree that the shock-value of some of the ideas of the *Republic* is important. Here Burnyeat seems to accept that an argument in the *Republic* can be directed towards addressing a need of the interlocutor or reader. But why then does Burnyeat consider any theorizing that doesn’t point to practical outcomes as “idle day-dreaming” (300-2)? Why require Callipolis’ practical possibility for the effects that Socrates wishes to accomplish in his interlocutors? Besides, Plato seems not to share the same scholarly scruples at idle theorizing when he has Socrates state: “You are amusing. You are like a man who is afraid of the many in your not wanting to seem to command useless studies” (527d).

22 During the dialogue, Glaucon and Adeimantus are willing to go along with anything. Then again, they are also forthright about every detail in the founding of Callipolis, as if it were a task they will soon face. As Burnyeat writes: “The ideal city is built in our imagination by persuasive argument, in such a way that successful persuasion in the world of imagination guarantees the possibility of success in the actual world” (*Utopia*, 308). That’s certainly how Glaucon and Adeimantus feel for the duration of the conversation. However, just as right after the conversation they will see no immediate political path to take, so also their ready agreement to the fantastic elements in Callipolis does not necessarily represent the more sober, considered view they will have after the conversation. Thus my argument for Glaucon and Adeimantus’ discomfort with the fantastic elements in Callipolis directly parallels Strauss’s argument (as I understand it) that Glaucon and Adeimantus will lose their political motivations after the conversation, despite their action-oriented attitude in the discussion. “For Glaucon it is more than enough that he will remember for the rest of his days and perhaps transmit to others the many
grand and perplexing sights which Socrates has conjured for his benefit in that memorable night in the Piraeus” (*City*, 137).

23 It may be wondered why Socrates should want to discourage philosophical ambition. But Socrates is not thereby doing harm to their philosophical potential; Socrates is just showing them their place, demonstrating that they are not the philosophers which they think they are. This is a way of fulfilling the Socratic mission of helping people get rid of false beliefs. But it is especially important since Glaucon and Adeimantus have political aspirations: as Plato’s seventh epistle shows, he was most worried or disgusted by Dionysus’ presumption to be a philosopher: “And if he had already discovered or learned these doctrines [i.e., Platonic philosophy] and regarded them as fitted for educating a liberal mind, how – unless he is a very strange creature indeed – could he have so lightly brought ignominy upon their teacher and guardian?” (*Ep*. 345b-c).

24 Compare 527d-528a, where Socrates must remind Glaucon that the discussion should be for his own sake.

25 “…nobody is capable of providing you with the influence you crave, neither your guardian nor your relatives, nor anybody else except me” (Alc. 105e).

26 See G.R.F. Ferrari, “Comments on Charles Kahn on Prolepsis”, esp. pp.3-4 for a theory of reading the dialogues paraleptically – that the understanding of one dialogue may require knowledge of the conclusions or assumptions made in another dialogue.

27 “…Alcibiades – undoubtedly, as boy and youth, very close to Socrates – was regarded by many in Plato’s time (cf. Lys. 14.16f., 29f., 35-40) as a traitor guilty of inflicting great and deliberate harm on Athens, and Socrates was blamed for his ‘teaching’ of such a ‘pupil’” (*Symposium*, ed. Dover 164).

WORKS CITED

All quotations from Plato are all taken from the Hackett *Complete Works*, except those from the *Republic*, which are taken from Griffith’s translation.


Brown, Eric. “Justice and Compulsion for Plato’s Philosopher-Rulers”. In


ERIN BEEGHLY

Room to talk? 'Archaeology' and 'Experience' in Early Foucault

“We are the creators of wrecks; there is nothing in our minds that anyone will manage to set afloat again.”

~ André Breton
“Cachée dans la soumission aux règles d’une tâche et dans la régularité d’exigences objectives non choisies, il peut y avoir une érotisation d’histoire- une passion altérante et altérée, j’oserai dire: une rage d’aimer”

Michel de Certeau.

Michel Foucault was a philosopher who celebrated the ‘death of man’, scorned psychology, and crafted a historical method that excluded the reflective, meaning-giving subject. In ‘What is an Author’, he argued that the concept of ‘author’ implying a creative individual was dubious, claiming the ‘author’ was really a function of discourse.¹ In ‘Thought from Outside’, he anticipated an ‘anonymous form’ that would rob the Cartesian subject of simple identity, the ‘right to say I’, and disintegrate the illusion of interiority.² Given these qualms about and down right hostility towards the concepts of ‘man’, ‘subject’ ‘consciousness’, and ‘author’, I will claim using evidence from the philosopher’s ‘archeological’ period(before 1970) that it is tricky, but not impossible to justify giving a lecture course or writing a book on Foucault and to understand how Foucault can justify speaking about himself. It is tricky because the philosopher’s archaeological method assumes the primacy of discourse over subjectivity and tends to skip over meaning in the world, seeming to preclude any talk of individuals. However, as I will argue, one can find in the notion of ‘experience’ that appears in Foucault’s literary criticism plenty of room to talk of Foucault.

At first glance, it seems as if Foucault’s project in his early years would reject any talk of individuals. His ‘archaeological’ method analyzed relations between statements, skimming over meaning, in order to discover the conditions of possibility that allowed people, things, objects, and concepts to show up in a particular way at a particular time. This analysis excluded concepts like ‘man’ and ‘consciousness’ necessarily, a point that Foucault repeated continuously in Archaeology:

…discourse is not the manifestation of a thinking, knowing, speaking subject, but, on the contrary, a totality, in which the dispersion of the subject and his continuity within himself may be determined.³ The analysis of statements operates therefore without reference to a cogito (AK 122).

It(discursive practice) must not be confused with the expressive operation by which an individual
According to Foucault, it was discourse *that* created the ‘subject’ and not the subject *who* created discourse. He claimed his ‘archaeological’ method would free him from problems inherent in a traditional history of ideas and from the unstable repetition that he had diagnosed in the human sciences in *Order of Things*.

Where do these claims leave those wanting to speak about Foucault as a philosopher or hail his creativity? One answer is, pretty much nowhere. His ideas in *Archaeology of Knowledge* and *Order of Things* seem to justify other kinds of structural analysis, like those proposed by Claude Levi-Strauss, Roland Barthes, or Jacques Lacan, but certainly not scholarly attention to individuals. In fact, Foucault explicitly advocated structuralist research in Chapter 10 of his *Order of Things*, claiming that these approaches, with their focus on anonymous processes and rules, offered an escape out of an epistemologically unstable situation. It seems then, if one would take Foucault’s advice and work seriously, instead of putting time and energy into lecture courses or books on him or his work, one would do ‘archaeology’ or engage in some other ‘structural’ enterprise, helping along ‘the death of man’, and necessarily abstracting individuals.

On the other hand, it may be possible to do a ‘structural analysis of knowledge’ while still speaking of Foucault. ‘Archaeology’ does not deny that ‘subjects’ or ‘authors’ exist nor does it refuse to the ‘archaeologist’ the right to use proper names. It rejects the idea that a stable knowledge of ‘man’ can be accessed through his subjective experience (a critique of the human sciences) and the idea that ‘man’, ‘subject’, ‘author’, etc are historically invariable concepts (a critique of naively done history and of ‘modern’ assumptions). It claims to offer a more stable kind of knowledge and to consider historical change. Foucault himself uses proper names of ‘authors’ in *Order of Things*, analyzing ideas associated with these ‘authors’, connecting them with others in the same *episteme* without any reference to individual subjectivity or intentionality; and in ‘What is an Author’, he allows for special attention to what he calls ‘discursive initiators’ (as opposed to great minds) like Marx, Freud, and Nietzsche whose texts are returned to time and again (WA 132). If one would present Foucault, not as example of a unified, a-historical individual, but as a ‘discursive initiator’ or as an ‘author-function’ (author as a function of discourse), then writing a book or lecturing on him could be done without issue.

Of course, one could skip justifying Foucault on his own terms if one
could prove that his ‘archaeological’ analysis did not present a more stable alternative to the human sciences as he claimed. If his method did not, in fact, dispel ‘all the chimeras of the new humanisms’ but replaced them with new ones, namely ones relating to discourse, then why play by his rules? What better alternative has he offered to anthropology? The analyses of Dreyfus and Rabinow and Beatrice Han all argue this point, claiming that Foucault’s method is logically unclear, lacks systematic cohesion, and that it falls prey to the doubles it claims to avoid, only in a different way than the human sciences. Even more pressing, they claim, is the fact that archaeology, unlike the human sciences, does not offer a coherent account of how discourse relates to humans. Foucault’s method focuses on discursive formations and ignores the social practices that support them, brackets (and hence does not account for) meaning in statements and the world, and is unclear on whether discursive formations function as rules in the minds of performers or if they are merely posteriori descriptions (BSH 87). A logical conclusion would be that, “...there is no place in archaeology for a discourse with social significance, no reason anyone should listen, and in spite of Foucault’s posturing, no reason anyone should write.”

In this light, it would be worth writing a book on Foucault (to refute his claims), but surely it would not do to devote a whole lecture course to him, for what positive contribution does he have to make? On another level, Dreyfus and Rabinow’s reading suggests that one can justify writing or speaking about Foucault however one pleases. Though he has convincingly argued that knowledge produced by the human sciences is unstable and ambiguous, so is the knowledge produced by his ‘archaeology’. Refer to Foucault as an ‘author-function’ or as a creative individual, what does it matter? But one need not refute Foucault to find justification for speaking of the philosopher in un-archaeological ways. Foucault acknowledges in numerous places that ‘archaeology’ is not the only legitimate kind of analysis. He does not want to say that it is a totalizing method, and in Archaeology makes sure to clarify ‘it[his method] is not a way of saying everyone else is wrong...trying to reduce others to silence’ (AK 17). He claims to allow(one may argue how sincerely or not) for the enterprises of a traditional history of ideas. In ‘What is an Author’, he begins the presentation, ‘For the purposes of this paper, I will set aside a socio-historical analysis of the author as an individual and the numerous questions that deserve attention in this context’ (WA 115). This is not a categorical rejection of socio-historical analyses(tied to the human sciences and involving ‘Marxist’ interpretation), but a presentation of his case, on its own merit, in a completely different field. Critical enterprises of all sorts are welcomed by Foucault.
But what of Foucault himself? Can he refer to himself as anything but an author-function, as controlled by language via discursive formations? If he does speak of his own ‘consciousness’ or subjectivity, does not this seem paradoxical? What status does he give his own meaning giving activity, his speech?

Here, the difference between how Foucault logically can and how he does justify speaking of himself in practice is a big one. If we are to believe Dreyfus and Rabinow, Foucault as an archaeologist can have no incentive to speak (as his discourse is not privileged with meaning and can serve no critical function) and, in theory, no ability to speak as an individual (for the discursive formations are speaking for him). Logically, all this seems to follow from the positions he puts forth in the *Archaeology of Knowledge* and *Order of Things*. However, the fact remains that Foucault does justify speaking of himself as a philosopher and writer in numerous articles that Dreyfus and Rabinow do not mention. These articles complicate his archaeological program even more because they include talk of play in and behind the structures of language, of disunity and dispersion in the self.

In Foucault’s literary criticism including ‘Language from the Outside’, ‘What is an Author’, and ‘Preface to a Transgression’, the philosopher explains how he can justify and, moreover, delight in speaking of himself. Significantly, his justifications fall outside the ‘archaeological’ method as he speaks of ‘experience’, a term with a secure place in the Foucault oeuvre from the beginning, but that lies low in *Archaeology* and *Order of Things*. What he has in mind by the term ‘experience’ is naturally ambiguous and hails to early-mid 20th century French literature (Blanchot, Breton, Rousseau, Bataille, etc.) and the idea that language can be an ambiguous space in which one can hide and play, where the words one do not refer back to an imperial, unified self but to a lack, a dispersion of self (TO 12-13).

The act of speaking, of writing, then, for Foucault is an experience of language, an experience that convinces one of one’s own fragmentation. He says this much in ‘What is an Author’ when he describes writing as, “a game that inevitably moves beyond its own rules and finally leaves them behind...an opening where the writing subject endlessly disappears” (WA 116), and in ‘Preface to a Transgression’, where he delights, “even the philosopher does not inhabit the whole of his language like a secret and perfectly fluent God.” With language as a game and a dispersion of self, he is justified in speaking of himself and his work; he is not pretending to be any kind of imperial ‘subject’. And more, freedom seems immanent in this play of ‘transgression and limits’ as the experience of language allows one to transcend discursive determinations. The need for this ‘experience’
is more than justification for Foucault to write.

What are the specifics of the discursive relations/‘experience’ dynamic? How do they relate? Are the ideas compatible or inconsistent? These are interesting and important questions, but well beyond the scope of this paper. The important observation for my purposes is that, because of these mixing ideas in his early work, one can find many justifications to write about the philosopher, lecture on him, and also understand why Foucault is able to speak of himself. On the other hand, given his arguments, perhaps one should say nothing. It could be that Foucault realized this last point, as a six year ‘silence’ followed *Archaeology*, a time during which the philosopher modified his methodology to include a new tool, genealogy, that would explain how statements were selected and allow for a more effective social critique. Hence the shift from discursive relations to the study of institutions and the focus on power relations of Foucault’s ‘middle period’.

Throughout the changes, however, Foucault preserved his qualms about the concepts of ‘subject’, ‘author’, ‘consciousness’, and ‘man’, continuing to argue that critical history should not rest on such assumptions. And so related question of, ‘on what basis can you talk’ persisted, criticism moving along similar lines as in his early period. Whether it was discursive or power relations in control, it seemed that individuals were forgotten. But as I have argued, the Foucauldian picture was never that simple, nor would it be. As the philosopher pointed out in his later years (when he began to write on ethics), he was always interested in the question of the ‘subject’.\textsuperscript{13} This was a fragmented and diffuse subject, one who had constraints that needed accounted for, but one that nevertheless possessed a capacity for freedom and self-reflection,\textsuperscript{14} a capacity unlocked by speaking and writing, and for Foucault himself by the writing of history.

NOTES

\textsuperscript{1} Foucault, Michel, ‘What is an Author,’ in *Language, Counter-Memory, Practice: Selected Interviews and Essays by Michel Foucault*. Trans. Donald F. Bouchard. (Cornell University Press 1977), 138-148.
\textsuperscript{2} Foucault, Michel, ‘Thought from the Outside,’ in *Foucault/Blanchot*. Trans. Brian Massumi. (Gallimard 1994), 47.
\textsuperscript{3} Foucault, Michel, *Archaeology of Knowledge*. trans. A.M. Sheridan Smith.
As it turns out, Foucault does not have a consistent story about how discourse creates subjects, nor does he pin down the exact relation between these ‘anonymous rules’ and speech acts of individuals. I will explain this ‘problem’ briefly when discussing Dreyfus and Rabinow’s critique of archaeology. What I want to emphasize here is merely that Foucault wants to skip over subjectivity, intentionality, etc. in his account of language, favoring instead a quasi-structuralist approach that emphasizes anonymous processes and/or rules.


Dreyfus and Rabinow. Beyond Structuralism and Hermeneutics: Michel Foucault. (University of Chicago 1982), 90-100. Pg 98: ‘It is not surprise that archaeology, by thus both affirming and denying the finitude of its own discourse, turns out to be as unstable as its precursors”. For a slightly different, but parallel account of how Foucault repeats the doubles (cogito/unthought, return/retreat of origin, transcendental/empirical) see, Han, Beatrice. Trans. Edward Pile, Foucault’s Critical Project: Between the Historical and the Transcendental. (Atopia 2002), 60-69.

Ironically, Foucault also questions whether his writing has a critical function and whether he should write. However, he cites the bourgeoisie’s appropriation of writing and not the logic of his method as the cause of this loss. See footnote in DEI, ‘Qu’est-ce qu’un auteur’ 839 and ‘Folie, Littérature, Société’ 983. I have been able to find neither of the footnote in the first article, nor the entire second article in translation.

Indeed, it seems the later work of Foucault (especially that on ethics) justifies those who give him attention and not the work he produced in his ‘archaeological’ period.

Indeed, Foucault collaborated with Dreyfus and Rabinow and even wrote extra essays to supplement their book.

Early evidence of this theme is found in Foucault’s Mental Illness and Pyschology (1954,62), Madness and Civilization, and in numerous interviews contemporaneous with Order and Archaeology. For example: ‘A Swimmer Between Two Worlds in Essential Works’ Volume 2 546-567.

Why does language convinces one of ‘one’s own fragmentation’ according to him? What theoretical underpinnings accompany this account of language and experience? I hope to answer this question in a Fall ‘03 project on the notion of ‘limit experience’ in early Foucault. Currently, I make only tentative and general comments on what he may mean by the term ‘experience’.


In the forward to Foucault’s Mental Illness and Psychology, Dreyfus argues for this interpretation of Foucault. I agree with him. See, Foucault, Michel. Mental Illness and Psychology. trans Alan Sheridan. (University of California 1987), xxxi.
MITCHELL ANDERSON

The Perceptual ‘Mirror’ in Merleau-Ponty’s Phenomenology

“A child said, 'What is the grass?’ Fetching it to me with full hands; How could I answer the child?...I do not know what it is any more than he.”

~Walt Whitman, Leaves of Grass
In *The Phenomenology of Perception* Merleau-Ponty seeks to overcome what he considers to be the fundamental problems of empiricism and intellectualism by taking a phenomenological approach to perception. His phenomenological account is an attempt to describe the pre-reflective, embodied experience of perception, which he thinks is overlooked by empiricism and intellectualism. Thus said, the following essay has a three-fold purpose. Firstly, it seeks to situate the uninitiated reader of Merleau-Ponty in a philosophical context, through which a more comprehensive understanding and appreciation of Merleau-Ponty’s phenomenology can be attained. Secondly, it seeks to elucidate Merleau-Ponty’s positive description of our experience of the world. In doing so, it will focus on the following metaphor which he uses to describe perception: “every object is the mirror of all others” (Merleau-Ponty, 68). It is my argument that his reason for conceiving of perception in this manner is twofold: primarily it is to account for the descriptive facts of the phenomenon of perception, but also it is to give an account of the natural crypto-mechanism which runs through perception and leads to the constitution of stable objects and the mental positing of a determinate world. Although at this point the notion of a crypto-mechanism coursing through perception is admittedly vague, throughout the essay it will become clearer. Finally, I shall conclude the essay by arguing that the metaphor which Merleau-Ponty uses—that all objects are the mirror of all others—is crucial to Merleau-Ponty’s phenomenological description of perception in that it allows his positive description of perception to account for the failures of empiricism and intellectualism.

To begin, it seems appropriate to juxtapose Merleau-Ponty’s phenomenological account of perception with empiricism and intellectualism so that in the latter part of the essay the groundwork has already been laid for showing both the failures of empiricism and intellectualism, as well as how Merleau-Ponty can account for the cause of those failures in his description of the phenomenon of perception. As M.C. Dillon has demonstrated in *Merleau-Ponty’s Ontology*, perhaps the most instructive way to present Merleau-Ponty’s phenomenology to the uninitiated reader is to consider it an attempt at resolving Meno’s Paradox. Briefly, Meno’s Paradox comes from a discussion between Meno and Socrates in which Meno poses the following dilemma: “But how will you look for something when you don’t in the least know what it is? How on earth are you going to set up something you don’t know as the object of your search? To put it another way, even if you come right up against it, how will you know that what you found is the thing you didn’t know?” (Plato’s Meno). According to Merleau-Ponty both empiricism and intellectualism fail to resolve this paradox. As he says:
Empiricism cannot see that we need to know what we are looking for, otherwise we would not be looking for it, and intellectualism fails to see that we need to be ignorant of what we are looking for, or equally again we should not be searching (Merleau-Ponty, 28).

Although the remainder of this essay shall not explicitly endeavor to show how Merleau-Ponty resolves Meno’s Paradox, though implicitly it will become evident, it seems appropriate here to briefly comment on how it is that empiricism and intellectualism fail to do so. To begin, contrary to phenomenology, both empiricism and intellectualism are rooted in the Cartesian tradition of modern philosophy which believes that there is a split between the subject (consciousness) and the object (things outside of consciousness). This gap between the subject and the object on Merleau-Ponty’s account can be understood as one of the assumptions that contributes to the failures of both philosophies. To Merleau-Ponty, this assumption is rooted in the failure of empiricism and intellectualism to recognize that the body—the bearer of our existence and our way of being in the world—is what makes perception possible. Merleau-Ponty’s notion of the body will be expounded later, and shall prove essential to understanding his account of how perception takes place. To continue, empiricism asserts that the transcendent world—the world outside of experience—acts upon consciousness and is the only means by which we can come to know anything. By positing an objective world that exists outside of experience, while at the same time claiming that all knowledge is derived from experience of this transcendent world, it seems that empiricism, along with not being able to overcome Meno’s Paradox, also fundamentally passes over the phenomenon of perception. Simply put, if our only source of knowledge is a world that exists outside of experience, how can we be certain that in experiencing this world we are coming to have any knowledge of it all?

Intellectualism also fails to resolve Meno’s Paradox. While empiricism claims that all knowledge is derived from experience of the transcendent world, intellectualism holds that all knowledge is a priori. In other words, all knowledge is known by the subject prior to experience. Rather than the transcendent world impinging upon consciousness, which is what the empiricist believes, the intellectualist claims that consciousness constitutes our experience of objects in the outside world. It seems then that intellectualism fails to resolve Meno’s Paradox because if all knowledge is known a priori, what is the point of searching for knowledge, since we already know what
we are searching for? Put differently, intellectualism rejects the notion that knowledge can be acquired through experience and in doing so renders pointless all endeavors of the human mind that look to the outside world for answers, such as the sciences and even philosophy.

With an understanding of how both empiricism and intellectualism fail to resolve Meno’s paradox, it now seems appropriate to begin a discussion of Merleau-Ponty’s positive account of the phenomenon of perception. To begin, according to Merleau-Ponty, we, as perceivers, are not objects located in the brute, objective space of the empiricist and the intellectualist, but rather we are embedded in ‘bodily space’. That is, we are embedded in a spatio-temporal reality which is always understood in relation to how our bodies are situated in it. Furthermore, according to Merleau-Ponty, our bodies are not machines as Descartes had thought, nor can they be understood as solely consisting in parts; rather our bodies must be understood as they are lived. This is Merleau-Ponty’s idea of the ‘lived body’ and it is crucial to understanding his account of perception. In effect, our body must not be understood as Descartes had believed, namely as an object in the world; but rather our body must be understood as the bearer of our existence in the world, as the way we inhabit the world and are aware of it. We as subjects are intertwined in a perceptual world and it is only because we are embodied beings that we can open onto it and have objects disclose themselves to us. To Merleau-Ponty we perceive objects with our bodies, and it is only in virtue of our having bodies that perception is possible.

With this understood, I shall now endeavor to explain how it is that perception takes place. According to Merleau-Ponty we can never experience things apart from our own bodily perspectives on them. The structure of perception is such that an object cannot disclose itself for me without concealing another of its aspects. Moreover, the very same object which conceals some of its properties from me is never perceived in absolute isolation; but rather it is perceived in so far as it is part of a perceptual structure in which it cannot reveal itself without concealing other objects in the system. Merleau-Ponty refers to the perceptual structure in which objects are distinguished and disclosed to the perceiver as the “object-horizon structure.” It is in virtue of the object-horizon structure that perception has the continuity that it does. That is, the world is not given to us in perception as a series of frames. It is not given to us as a discontinuous and fragmented motion picture. Rather, perception has a certain continuity in that as one object becomes the center of perception, all the other objects of the system recede away to the periphery, while still remaining present to me. As an embodied being, I always have the impending power to shift my glance to various
objects in the perceptual field, while retaining the other objects, currently
not under the scrutiny of my gaze, as potential candidates for scrutiny.

In order to further understand this idea of how perception is always
embedded in an object-horizon structure, it shall prove instructive to discuss
Merleau-Ponty’s example of how perception takes place in a film. Through a
discussion of how perception occurs in a film, the notion of the object-horizon
structure becomes clearer. Throughout a scene in a film the camera focuses
on many objects, and upon a return to a certain object, e.g. an ashtray, the
perceiver does not identify it in the strictly perceptual sense as being an object
which was formerly on the horizon but now has drawn forth his gaze and
become disclosed. Rather the perceiver remembers the ‘ashtray’ intellectu-
ally as being a determinate object that always existed, yet was not present
in previous perception. That is, upon a shift in the angle of the camera, the
ashtray vanishes, and upon a subsequent return to the ashtray, it reappears
abruptly. Thus in a film the identity of the ashtray is maintained only by the
perceiver’s ability to remember that it was supposed to be there the whole
time. In normal perception this is not what occurs. Rather, my body presents
me with a particular perspective on the ashtray, and when I shift my glance
away from it, it still remains present to my perception as an “abode open to
my gaze” on the horizon (Merleau-Ponty, 68). Thus, in normal perception
we don’t just see what is on the retina, we experience ourselves as opening
up the “horizon”, as exploring what was already marginally there in the
background, or what was already perceptually implied by the whole of the
system.

It now seems appropriate to begin a discussion about why it is that
Merleau-Ponty claims that in perception we perceive objects as if “all ob-
jects were the mirror of all others” (Merleau-Ponty, 68). To begin, as was
discussed earlier, perception is possible in virtue of the fact that I have a
body that situates me in the world and presents me with a perspective on
my surroundings. By no means, however, is my perception limited to my
singular perspective. If this were the case perception would be superficial
and its richness would be lost. Consider the following scenario: I am sitting
in my room and my body is situated on the couch in such a way that I have a
perspective on my surroundings. This seems self-evident but is nonetheless
crucial. My body is my mode of communicating with the world and as I
perceive my guitar and my stereo, as well as all of the other objects currently
in my perceptual field, they disclose themselves relative to the positioning
of my body. However, I am not stuck here on the couch, and it is in virtue
of this fact, namely that my body provides me with the ability to manipulate
my surroundings, that I can have perceptual “implications.” For example,
I do not just perceive my guitar as if it only consisted in the six strings and the fret board currently present to my perception, nor do I judge that the back of the guitar exists, but rather as Merleau-Ponty says “[I] apprehend an imminent sense in the sensible before judgment begins” (Merleau-Ponty, 35). That is, by virtue of being bodily engaged in the world, I sense that my guitar is more than just these six strings that I currently can see. It is not reason that tells me this. I do not need to judge. Rather, while my perspective limits my perception of the guitar in one sense, it also extends to me the promise, or the implication, that from elsewhere the whole of my guitar could be seen. On Merleau-Ponty’s account, in perception I perceive all aspects of the present object under scrutiny (e.g. the guitar) as if it were being seen by all the other objects currently on the horizon—as if all objects were the mirror of all others.

In order to more thoroughly understand this notion that all objects are the mirror of all others in perception, it will be helpful to reiterate the intimate relationship between the body and the world in the production of perceptual experience. This will leave a twofold purpose. Primarily it will help us to understand Merleau-Ponty’s account of the descriptive facts of perceptual experience. An interesting point will also emerge which will prove instrumental in our later discussion of how the ‘crypto-mechanism’ naturally distorts the primal phenomenon of perception.

As has already been discussed, according to Merleau-Ponty we are embedded in ‘bodily space’ and perception occurs in an object-horizon structure. However, what has not been discussed is that on Merleau-Ponty’s account we have a certain body disposition, or body readiness which affects the way objects looks as well. This is not to say that my body disposition is the result of an intellectual process, which subsequently affects perception. Rather, to Merleau-Ponty, it is only because we have bodies that are involved in coping in this world, that have anticipations, and that are able to shift and move and manipulate their surroundings, that we have certain perceptual implications. Let us now take Dreyfus’ example of the man walking on Telegraph Avenue to reveal the intuition behind this idea. Suppose that there is the figure of a human being walking down Telegraph Avenue who from behind appears to have the form of a woman, the long hair of a woman, and the apparel of a woman. We consequently form a set of anticipations relative to our perception of this effeminate figure. That is, our body disposition is affected by our perception of this effeminate figure. We sense that this is a woman in front of us. From behind this body what is directly present to our vision is nothing more than the back of the body. However, the body has a front side, and we experience this as both indeterminate as well as something that we can find
out more about. The fact that our body endows us with this possibility to explore further the present object being perceived is what allows us to see the object in such a way that its hidden aspects are perceptually implied by all the other objects in our perceptual field. Thus, in actually perceiving the back of the body, “I attribute to it not only those qualities visible from where I am” but also all of those attributes in which the cement, the shopping sign overhead, as well as the person passing by can ‘see’. That is, I perceive the back of the body as if it were a woman, as if her concealed aspects (i.e. face and eyes) were effeminate in nature. As Merleau-Ponty says, “any seeing of an object by me is instantaneously reiterated among all those objects in the world which are apprehended as co-existent, because each of them is all that the others ‘see’ of it” (Merleau-Ponty, 68). That is, I perceive the back of the body as if all objects were the mirror of all others. Now if I were to find out that this effeminate figure was in fact a man, my perception of the back of his body would be altered. My body disposition would change in relation to my perception of the back of the body. In a word, the way that my body disposition configures itself is in such a way that it is affected by my perception of an object, yet in turn, the way it configures itself affects my perception of that self same object.

Thus, it is the ‘mirroring’ aspect of perception, (i.e. perceptual implications) which influences my body disposition in relation to the object being perceived, and in turn influences the way in which the perceived object looks to me. At this point, I will briefly discuss how it is that the ‘mirroring’ feature of perception induces a certain stability in perceptual experience. According to Merleau-Ponty, in perceptual experience the front of the body is concealed by the back of the body, though present to the shopping sign as well as to the person passing by. As I explore the “inner horizon” of the back of the body, all of these other objects in the system, though on the periphery of my vision, present their aspects to my perception of the back of the body, and in doing so, “guarantee the identity of the object throughout the exploration” (Merleau-Ponty, 68). That is, I do not just see the present object under scrutiny from my singular perspective, but rather the whole of the object—its hidden aspects as well as those that I presently see—is perceptually implied by all other objects in the perceptual structure. It is in this sense that the identity of the object is guaranteed by all of the other objects currently on the horizon of my vision. This must not be misinterpreted. The identity of the object is not guaranteed in the objective sense, as if it were some completely determinate thing. Rather, its identity is guaranteed in virtue of its being in my perceptual field, in virtue of its being in my domain where its identity is already perceptually implied by the whole
of the system. It is in this facet of perception, in which the objects of the perceptual structure mirror each other and assure the identity of the present object under scrutiny, where perceptual experience attains a certain aspect of stability.

The question now on our hands is what to make of this seemingly strange notion that perceptual experience, though fundamentally indeterminate, culminates in stability as a consequence of perceptual implication. In order to understand the significance of this, an explication of the crypto-mechanism seems necessary, if not already overdue. Here I will adopt Dreyfus’ view that the crypto-mechanism consists in two parts: first, perception hides itself, and second, that thinking actively extrapolates from the phenomena and somehow retroactively covers up perception. Intuitively, the crypto-mechanism can be understood in Merleau-Ponty’s phenomenology as an illusion inherent in our embodied perception of the world that leads us to posit a world of determinate objects while also taking us away from our phenomenal experience, in which the perceptual world is fundamentally indeterminate.

My claim is that the aspect of perception, in which all objects mirror each other and thus induces some sort of stability in the phenomenon, is actually the last stage of perception hiding itself and the incipient stage of thinking retroactively covering perception up. In this sense it serves as the integral link, facilitating the natural progression of the ‘crypto-mechanism’ in its constitution of a completely determinate world. Once I have demonstrated this, I shall engage in the final endeavor of the essay, which is to show how it is that this integral link—the aspect in which all objects mirror all others—is crucial to Merleau-Ponty’s phenomenology in that it allows him to explain, within his positive account of the phenomenon of perception, the failures of empiricism and intellectualism.

To adequately engage in this task however, we must first come to terms with the dual nature of the ‘crypto-mechanism’ and understand that these aspects are not contradictory but are rather a natural progression which culminates in the transcendence of actual perceptual experience and the positing of an objective world. From here, we will then be able to understand how it is that they both are linked at the one integral stage of perception in which objects act as the mirror of all others. According to Merleau-Ponty the perceptual world has connections of motivations between its aspects, and it is in this way that the perceptual field motivates and organizes itself so as to enable a perceptual experience of stable objects. Drawing from Gestalt Psychology, Merleau-Ponty holds that the perceptual field always tends towards equilibrium and stability. It is through this organization of the objects present in perception and the instantaneous workings of the
background, which take into account lighting, shade, and color, etc, that the instability of the primal perception is hidden from perception itself. To put it differently, the perceptual field has a way of taking account of all of the details of perception and organizing them to produce a stable scene. Perception hides itself in the sense that it pulls the perceiver’s attention away from the background and the field and to the stable objects. For example, in a lighted room a perceiver sees all of the objects as illuminated objects. What the perceiver does not see is the perceptual field allowing the objects to be seen as illuminated, by organizing the scene in such a way that the actual light withdraws, creating an aspect of stability. Merleau-Ponty’s example of a ship run aground furthers this point. On a beach there is a ship which has run aground, and its mast extends into the forest bordering the beach. Upon initial perception of this scene, the mast and the forest are indistinguishable, and in this sense the whole scene is somewhat indeterminate. However, according to Merleau-Ponty, “there will be a moment when these details [of the ship] suddenly become part of the ship, and indissolubly fused with it” (Merleau-Ponty, 17). It is in this moment that the background and the perceptual field motivate their aspects to produce a stable scene which covers up the initial indeterminacy of the perception.

From this story of perception hiding itself naturally follows the sequel, whereby thought actively extrapolates from the phenomenon and then retroactively posits the determinate existence of an objective world. As seen by the workings of the perceptual field, our experience always tends towards more stability and more plenitude in the perceptual object. What occurs as a natural result of this tendency towards stability in the perceptual field is that consciousness, “forgetful of the perspectivism” of experience, deduces the origins of perceptual experiences from relations among objective objects. With the instability of perception blinded by perception itself, thought is able to extract the notion of stability from perception and thus posit that an objective world has always existed and is the source of all experience. As Merleau-Ponty states, “the positing of the object therefore makes us go beyond the limits of our actual experiences” and into a world of already constituted, autonomous beings.

To understand my claim, that the aspect of perception by which objects mirror each other is the final stage in perception hiding itself and the incipient stage in the positing of an objective world, we can now return to the example of Telegraph Avenue. As discussed earlier, upon perceiving the back of the person walking down Telegraph, we have a set of anticipations which influence the way it looks. Our set of anticipations, though fundamentally indeterminate, have a claim to determinacy in the sense that they are determi-
nate enough to be wrong. That is, we attribute to the front of the effeminate figure that which we have apprehended from the back. What gives our set of anticipations a claim to determinacy is that the identity of this body is guaranteed by the system of objects which are bound together in the scene. The cement below the body, the shopping sign above the body, as well as the person passing by from the other side, draw forth a determinacy in my set of anticipations, in that I experience the possibility of exploring further and perceiving from another angle what I currently see from the back. That is, the current object of my present vision is tending towards stability as a result of the inner workings of the perceptual field, yet it is only in virtue of my being bodily engaged in the world that I can see the object as a stable, already constituted thing. More specifically, I attribute in my perception of the object all of those aspects which are perceptually implied by all of the other objects that surround it, and consequently, my perception of the object is no longer ambiguous as it once was. My perception of it is no longer the perception of an indeterminate thing, as was the ship and the forest; but rather it is the perception of a determinate existence, whose features are constituted and available for further exploration. As Merleau-Ponty says, “I see an object in so far as objects form a system or a world, and in so far as each one treat the others round it as spectators of its hidden aspects and as guarantee of the permanence of those aspects” (Merleau-Ponty, 68). The workings of the perceptual field organize the details of the scene so as to produce a sense of stability and equilibrium, however, it is the ‘mirroring’ aspect of perceptual experience, the fact that all objects perceptually implicate each other, which endows our perception of objects with an even richer level of stability and determinacy.

As a result of perception hiding itself, whereby the mutual implication of objects is the last and decisive stage, the second aspect of the ‘crypto-mechanism’ is facilitated in its endeavor of positing a determinate world. We, as perceivers, do not posit an objective world directly as a result of the internal workings of the perceptual field, but rather it is a direct result of the objects in the system guaranteeing the identity of the object being perceived, which leads to our positing of an objective world. That is, it is a direct result of the ‘mirror’ aspect of perception. Once the scene is stabilized by the perceptual field and the object being perceived is ‘established’ through perceptual implication, we believe that our experience is being produced by a projection of external impressions on our retina. It is here, in the formation of a belief about the world, where “I detach myself from my experience and pass to the idea” (Merleau-Ponty, 71). In going beyond actual perceptual experience and the synthesis of horizons, “objective thought is formed,”
which finally causes us to lose contact with perceptual experience, of which it is the outcome and the natural sequel” (Merleau-Ponty, 71).

To conclude, according to Merleau-Ponty it is in virtue of my having a body, which always situates me in the world and presents me with a perspective on it, that perception is possible. Objects could not disclose themselves to me otherwise than in perspective, yet at the same time my particular perspective does not limit my perception, but rather it opens up the possibility of shifting my perspective and perceiving objects as if all objects were the mirror of all others. Ironically, it seems that on Merleau-Ponty’s account it is this aspect of perception, in which all objects mirror each other, that leads us away from the subjectivity and indeterminacy of our phenomenal experience into a perception of objects with seemingly stable and determinate properties. That is, although my body is anterior to all thought and is my primordial existence in the world and my mode of communicating with it, it is only because I have a body that I perceive all objects as the mirror of all others. In this regard, it seems that it is only in virtue of my being bodily engaged in the world that perception is always tending towards stability, and that perception inevitably culminates as the “perception of something” from somewhere (Merleau-Ponty, 70). It can thus be concluded here that although I perceive objects with my body, and in this sense perception is always perspectival, it also seems that according to Merleau-Ponty, my body is what allows for perceptual implications; and in this regard, it is only in virtue of my having a body that the crypto-mechanism can lead me into positing the existence of a determinate world. It can thus be seen that herein lay the failures of empiricism and intellectualism. As a consequence of the stability brought to our experience through what Merleau-Ponty terms the crypto-mechanism, the empiricist and the intellectualist have deduced the origins of perceptual experience from relations among objective objects. That is, they have failed to recognize the perspective and the subjectivity of our phenomenal experience. In a word they have objectified the body; they have conceptually treated it as just an external object among other objects. The only difference between the failure of empiricism and intellectualism is that the empiricist believes that perception is the result of external objects impinging on my retina, whereas the intellectualist believes that “I am producing the perceived perspective by the projection of the objects on my retina” (Merleau-Ponty, 71). Thus to Merleau-Ponty, the world that we inhabit is a “world [that] is always ‘already there’ before reflection begins” and thus we must return to our phenomenal experience of it in order to describe “the emergence of being”. In order to describe the phenomenon of perception, we must delve beneath thought and return to our primordial...
embodied existence, which is anterior to thought.

WORKS CITED


"I have no color prejudices nor creed prejudices. All I care to know is that a man is a human being, and that is enough for me; he can't be any worse."

~ Mark Twain

"Twas in another lifetime, one of toil and blood
When blackness was virtue, and the road was full of mud
I came in from the wilderness a creature void of form
'Come in', she said
'I'll give you shelter from the storm.'"

~ Bob Dylan
Introduction

In *Determination of the Concept of a Human Race*, Kant first introduced the question of race. He described four different races within the human species, distinguished only by gross morphology, which was supposed to be indicative of a more fundamental distinction, an inherited racial essence. Racial essences have since turned out to be a myth. However, if asked to do so, each of us could easily identify ourselves as belonging to a specific race, or as a mixture of several. What do we mean when we describe someone as belonging to a particular race, given the fact that biologically speaking, there is only one human race? Are we only referring to a loosely defined set of physical characteristics? Or is there something more significant picked out by the different categories Kant described? In light of the biological evidence against Kant’s conception, how should we conceive of race?

The purpose of this paper is to introduce the question, “how should we conceive of race?” into the affirmative action debate. In the first part of this paper, I will lay out an argument put forth by Ronald Dworkin in defense of affirmative action, which claims that affirmative action does not violate the rights of white males. In the second part, I will briefly assess a weakness in his argument, namely, his assumption that objections to race-based classifications only apply when one is excluded because of contempt for one’s race. I will then lay out a counter-argument put forward by Antonin Scalia that I believe challenges this assumption. Scalia assumes that if we can show that a system of race-based classification is not motivated out of contempt for a particular race, then objections to this system no longer hold any water. In the third part I will suggest that Scalia’s objection conflates two notions of race, one biological and the other socially constructed. I will characterize these two ways of thinking about race, support the view that the biological notion of race is a myth, and claim that Scalia’s objection to affirmative action is mostly directed at a biological notion of race. In the final section I will examine what Scalia and Dworkin’s arguments mean given the absence of biological races. I will suggest a way to strengthen the weakness in Dworkin’s argument, and I will argue that Scalia’s objection does not apply to affirmative action as we should conceive of it. Finally I will consider objections to the defense of affirmative action that I have proposed, and try to counter them.

I. The affirmative action debate: Does the white male have a fundamental right that is violated by affirmative action?
There are two different lines of argument that one frequently finds in defense of affirmative action. One strategy grants that white males have a right to an equal chance at a university position (student or faculty), but claims that this right can (or must) be overridden in favor of rights of others that are more stringent. An example of this line of argumentation can be found in Judith Jarvis Thompson’s article, “Preferential Hiring”, in which she argues that a policy of preferential hiring of blacks and women, two types of members who have been historically wronged by the community, is not a violation, but merely an overriding, of the rights of white males to an equal chance at a university job. For the purposes of this paper I am going to abandon discussion of this particular strategy, and focus solely on the second line of argument one finds, which claims that there is no fundamental right that is violated by affirmative action.

This second strategy of defense is taken up with respect to the Bakke case¹ by Ronald Dworkin in “Are Quotas Unfair?”. Dworkin’s argument begins with the claim that if the UC Davis quota policy violates Bakke’s constitutional rights, it would have to be because affirmative action violates some moral principle. There is no formal or technical law, he says, that expressly forbids affirmative action, nor does any reasonable interpretation of the Constitution, the Civil Rights Act, or any other congressional enactment support such a conclusion. Thus, Bakke must have an important moral right that affirmative action violates. But, asks Dworkin, what could that right be? He considers three possibilities: the right 1) to be judged on his merit, 2) to be judged as an individual, rather than part of a group, and 3) not to be excluded because of race alone.

Dworkin dismisses the first possible right (to be judged on his merit) because he thinks that there is no combination of abilities, skills or traits that constitutes merit in the abstract. In other words, admission into medical school is granted on contingent factors that happen to make one a better candidate to be a useful doctor in a given set of circumstances. If fast hands count as merit in the case of prospective surgeons, then they do so because they would enable one to better serve the public. Likewise, Dworkin argues, black skin might constitute merit because it is a socially useful trait in particular circumstances.² Thus, the right to be judged on merit is not actually an objection to the Davis quota policy, because race could be a legitimate means of determining merit in the circumstance where it would best serve the needs of the public.

Dworkin also dismisses the second right, (to be judged as an individual, rather than as a member of group) because colleges routinely dismiss applicants as part of a group, without assessing them as individuals beforehand.
Two examples of this are minimum GPA standards, and maximum age standards. In these cases, schools assume that applicants with even one point below the minimum GPA, or one year over the age of 30, will have less to offer to the medical profession than younger applicants with one point above the minimum GPA, even though individual treatment of the application or an interview with the applicant might have revealed otherwise. If a school were to make the presupposition that no white doctor can do as much to counteract racial imbalance in the medical profession as a well qualified and trained black person, he says, this presupposition is no less plausible that corresponding presuppositions about age or GPA. Thus, if the former two common admissions procedures do not violate a fundamental right to be judged as an individual, the latter would not violate any such right either.

The third right (not to be excluded because of race alone) sounds plausible, according to Dworkin, because it suggests the principle that every citizen has a constitutional right not to suffer disadvantage in competition for a public benefit because of prejudice or contempt for his race, religion, sect, or any other group he belongs to. The crucial point here is that exclusion based on race is considered unacceptable only if it is motivated by prejudice or contempt. It is obvious in Bakke’s case that he was not excluded because of prejudice or contempt for white males, who at that time (most likely) dominated the student body and the admissions committee at the UC Davis medical school. Dworkin points out that there are many factors about a person that she cannot control, such as intelligence, age, athletic ability, or region of origin, which may be legitimately used as grounds for denying admission. It is true that if Bakke had been black, he would have been granted admission. However, it is also true that he would have been granted admission if he was more intelligent. In both cases, says Dworkin, he was excluded by a rational calculation about the most socially beneficial use of limited resources for medical education. Thus, Bakke’s fundamental right was not violated, because he was not excluded because of prejudice or contempt for his race. Evidently, it still stands that the Davis policy is excluding people based on their race alone (even though this exclusion is not motivated by prejudice or contempt), which leaves Dworkin open to the objection that it is Bakke’s fundamental right not to be excluded just because of his race.

II. The objection (to Dworkin) that race-based classification is racist

The most contentious point in Dworkin’s argument is the claim that the
“right not to be excluded because of race alone” is only plausible insofar as it suggests the principle that one should not be excluded because of prejudice or contempt for her race. One might object to this inference, and say that, in fact, one has an important moral right not to be judged by race alone for any reason, be it in good faith or out of prejudice, because racial preference is always racial discrimination against the group that is not preferred. To make distinctions between two candidates based on race, as opposed to genuine individual worth and qualification, is fundamentally racist, one might claim, and is contrary to the principles of equal justice that we ought to be upholding. Antonin Scalia makes an objection along these lines in his response to the Bakke case, “The Disease as a Cure”.

Scalia claims that, because it discriminates on racial grounds and is premised on racial indebtedness, affirmative action is racist, and therefore highly objectionable. He attacks the concept of restorative justice that he thinks underlies affirmative action by way of a reductio ad absurdum. Blacks have not been the only victims of unjust discrimination, he suggests, as our country has also undergone periods of discrimination against other ethnic groups, such as Irish, German, and Italian immigrants. In order to make up for this inequity, he suggests we should establish a Restorative Justice Handicapping System, which would assign RJHS points at birth based on an individual’s ancestry. What we loosely call “Aryans” would obviously receive the highest number of points, but in order to accurately determine racial debt, we would have to further divide this group into subgroups. The Irish, having arrived later, would owe less of a debt than the Germans, who in turn would owe less than the English. Identification of debtor races is only the beginning, however, as one has to account for dilution of the bloodlines by establishing scores for those of mixed ancestry, for example, a half-Italian, half-Irish handicapping score. According to Scalia, this system is obviously untenable and offensive, and demonstrates the racism at work in the argument for restorative justice.

The Restorative Justice Handicapping System], and the racist concept of restorative justice of which it is merely the concrete expression, is fundamentally contrary to the principles that govern, or should govern, our society. I owe no man anything, nor he me, because of the blood that flows in our veins. To go down that road (or I should say, to return down that road), even behind a banner as gleaming as restorative justice, is to make a frightening mistake. This is not to say that I have no obligation to my
fellow citizens who are black. I assuredly do—not because of their race or because of any special debt that my bloodline owes to theirs, but because they have (many of them) special needs, and they are (all of them) my countrymen and (as I believe) my brothers. . . .I am not willing to prefer the son of a prosperous and well-educated black doctor or lawyer—solely because of his race—to the son of a recent refugee from Eastern Europe who is working as a manual laborer to get his family ahead. The affirmative action system now in place will produce the latter result because it is based upon concepts of racial indebtedness and racial entitlement rather than individual worth and individual need; that is to say, because it is racist.3

Scalia’s argument against affirmative action is directly aimed at the argument for restorative justice, but it exhibits an underlying attitude towards racial classification in general. He implies that race is not a morally relevant category, and thus, its use cannot be justified, even if it is not motivated by prejudice or contempt for a particular race.

I am calling attention to Scalia’s argument because I feel that it is an excellent articulation of a more nebulous feeling that many people who object to affirmative action have, a feeling that if racial classification was unjust in the case of the Jim Crow laws, it isn’t fair to take race into consideration in a competition for academic positions, no matter how noble our motivation. I think this objection is misguided, because it does not distinguish between two possible notions of race. Scalia’s argument seems to be directed at an essentialist biological notion of race, one that is dubious at best. The question of whether race exists, and if so, what it is, is one that is conspicuously absent for the affirmative action debate, and it is at this point that I would like to introduce it.

III. What do we mean by ‘race’?

As I mentioned before, the question of what race is does not arise within the context of discussions about affirmative action. It seems to be generally assumed that (1) there are racial groups, and one’s membership in a group is (in most cases) somewhat easily determined, and (2) that group membership proceeds on the basis of inheritance. It is somewhat surprising to discover these assumptions in the literature surrounding the affirmative
I contend that there is an important distinction between two different notions of race, one biological and one socially constructed, that needs to be made in the debates about affirmative action in order for us to be clear on whether or not racial classification violates an important moral right.

Historically, race has been taken to refer to a group of people who are biologically linked by inherited characteristics that constitute a “racial essence”. Race, as identified by Kant and Blumenbach, refers to groups denoted by phenotypical characteristics, such as skin color, hair type, and facial features. But the heritable characteristics that constitute a “racial essence” were also thought to be responsible for much more significant differences in character, intelligence, and culture. Academics seem to be in general agreement that modern biology has shown this hypothesis about racial essences to be utterly false, and that the only thing that race picks out is gross morphology. Appiah provides a fairly in-depth explanation of the lack of genetic evidence for races in chapter two of his book, *In My Father’s House*. After presenting the data, he concludes,

...nineteenth-century race science sought in a heritable racial essence an explanation of what its proponents took to be the observed phenomena of the differential distribution in human populations both of morphological and of psychological and social traits. What modern genetics shows is that there is no underlying racial essence.

In other words, there are only groups of people with similar trivial gross morphology, and consequently, classifying them in terms of races seems somewhat arbitrary. As Appiah notes, we might as well classify people according to whether or not they are redheaded and freckled.

The reference Scalia makes to “the blood that flows in our veins” and his characterization of restorative justice as requiring one “debtor race” that needs to make up for its past wrongs instantiates the view that people belong to fairly clearly delineated groups in virtue of common “blood”, which is to say, some shared biological essence. On his account of restorative justice, one group (in virtue of being white) owes a historical debt to another group (in virtue of being black), which suggests that there is some meaningful essence holding the group together over time. Scalia is right in claiming that this concept of restorative justice is racist, because it relies on a biological notion of racial essence. However, this biological notion of
race is not necessarily invoked in affirmative action, and if it is the case that we can understand racial classification as proceeding on grounds other than biological essences, Scalia’s objection will lose its import.

Although racial biology has turned out to be false, it is obvious that in our society racial classification proceeds without much difficulty. Given the lack of a meaningful biological distinction, what does it mean to affirm that one belongs to particular racial group? In her paper “You Mixed? Racial identity without racial biology”, Sally Haslanger sketches an account of what it means to belong to a racialized group.\(^9\) In the absence of biological races, according to Haslanger, race is the \textit{social meaning of the geographically marked body}. That is to say, there are (contextually variable) ways of reading the body in terms of gross morphology that, in the case of race, are presumed to be evidence of ancestry from a particular geographical area. These characteristics are, in themselves, fairly meaningless, but given our particular socio-political history, these ways of reading the body have become inextricably linked to either subordinate or privileged social positions. Haslanger gives the following helpful definition,

\begin{quote}
A group is \textit{racialized} iff its members are socially positioned as subordinate or privileged along some dimension (economic, political, legal, social, etc.), and the group is “marked” as a target for this treatment by observed or imagined bodily features presumed to be evidence of ancestral links to a certain geographical region.\(^{10}\)
\end{quote}

If and how a group is racialized is not absolutely determined, according to Haslanger, but is dependent on the social context. At the moment in the U.S., Whites, Blacks, Latinos, Asians, and Native Americans are racialized groups insofar as (1) they are defined in terms of physical features, which are associated with a geographical place of origin, and (2) membership to them functions socially as a basis for evaluation. However, her account allows for other groups that are not currently racialized in the U.S., but have been in the past and could be again, such as the Irish, Italians, and Germans.

This account makes sense with regard to group-based oppression, but an individual’s experience can vary greatly, claims Haslanger, depending on the context and the extent to which she is racialized. Most people are not passive victims of racialization, she points out, but are agents capable of undermining or collaborating with the process. We can say that a person S is a member of the Asian race if and only if Asians are a racialized group in S’s society, and S is regularly or for the most part treated as a member of
the Asian race. However, we might conclude that some people do not have a stable race, she says, given the variation in the ways in which they are treated. To accommodate the contextual racialization of individuals, says Haslanger, we can say:

\[ S \text{ functions as a member of a racial group } R \text{ in context } C \text{ iff } \]

i. S is observed or imagined in C to have certain bodily features presumed to be evidence of ancestral links to a certain geographical region (or regions) where the group R is thought to have originated;

ii. that S has these features marks S within the background ideology of C as someone who ought to occupy certain kinds of social positions that are in fact subordinate or privileged (and so motivates and justifies S’s occupying such a position); and

iii. the fact that S satisfies (i) and (ii) plays a role in S’s systematic subordination or privilege in C, i.e., \emph{along some dimension} S is systematically subordinated or privileged when in C, and satisfying (i) and (ii) plays (or would play) a role in that dimension of privilege or subordination.\textsuperscript{11}

In other words, I am White in virtue of the fact that my body is marked by particular skin coloring, hair type, physique, and facial features, which are read as having ancestral links to northwestern Europe, these features mark me as being someone who ought to occupy a privileged social position, and these conditions have played a role in my systematic privilege in the context of a post-WWII United States.

An interesting facet of this account is that while heredity plays some role in the attribution of race, it is not a sufficient condition for being a member of a racialized group. In other words, if a person had a grandmother born in Africa who was marked in such way that she would be considered Black, he would not be a member of the same racialized group unless he himself had characteristics that mark him as belonging to that group. So, having a recent or distant ancestor who was racially marked as Black has nothing to do with your race \emph{unless you bear similar physical markings}. It seems inevitable that if both your parents belong to the same racialized group, you will also be marked as belonging to that group. This may solve the apparent problem of people who are marked in such a way that they
are considered to be White, consider themselves to be White, but who may have been considered Black by the Louisiana courts because they had an ancestor who was Black.

Thus far, Haslanger has provided us with an account of race (or racialized groups) that gives a different signification to racial classification than has been historically conceived. Haslanger goes on to give an account of racial identity that takes into account both the cognitive and unconscious ways in which an individual’s racial identity is formed. This is a crucial portion of her account which might ultimately be necessary to give a complete account of race as it is socially constructed. However, I am going to leave Haslanger for the moment, and explore the ramifications of this distinction between two notions of races as it applies to the argument that racial classification is unjust.

IV. The implications of a culturally constructed notion of race for affirmative action

As you will recall, the weak point in Dworkin’s argument was his claim that the “right not to be excluded because of race alone” is only plausible insofar as it suggests the principle that one should not be excluded because of prejudice or contempt for her race. Scalia (and other objectors to Dworkin) claims that, in fact, any classification based on race is inescapably racist, and therefore, unjust, even if it is not motivated by contempt.

Let us re-examine Dworkin’s question: is there an important moral right to not be excluded because of race alone? It seems to me that the intuition underlying this claim does suggest a principle that one should not be denied public benefits because of prejudice or contempt for her race. But I think that there is something further we can say, namely, that racial classifications that proceed on an essentialist notion of race are both incorrect and unjust, whether they are motivated by contempt and prejudice towards one group, or by a noble cause like restorative justice. In other words, we think that it is wrong to classify people in terms of race, and then exclude or include them solely on this basis, in those cases where one assumes that race has some biological (and thus, essentialist) basis. We might say that what we find objectionable about the Jim Crow laws and the one-drop rule is that they relied on, and were motivated by, a notion of race that we think is both incorrect and morally unsound. Biologically speaking, there is only the human race, so it doesn’t make sense to address racialized groups as if they have something biologically meaningful binding them together over several
generations. Scalia’s objection that he “does not owe any man anything, nor he me, because of the blood flowing in our veins” expresses this intuition perfectly. Does this mean that affirmative action is operating under such biological notions of race, and thus violates an important moral right?

I propose that, if we adopt Haslanger’s socially constructed notion of a racialized group, we can understand why affirmative action is not a violation of this principle. On Haslanger’s view, an individual belongs to a racialized group in virtue of the way her body is marked, and the way in which those markings shape her present experience of systematic privilege or systematic subordination. In this respect, it does not make sense to justify affirmative action in terms of restorative justice for a racial debt, (as Scalia claims it is justified), because the way in which an individual is racialized is contextually variable, and proceeds on an individual basis. In other words, an individual who “belongs” to a particular racialized group does so in virtue of how she is marked and how those markings are interpreted in the present context, regardless of who her ancestors are, or the particular wrongs they either perpetrated or suffered. If affirmative action should not be motivated by restorative justice, what should it be motivated by?

Depending on how an individual is racially marked in the context of post-WWII America, her experience is significantly shaped by the status of her racialized group. Despite the fact that racial biological essences do not exist, the social and material realities that an individual encounters because of the ways in which her body is marked are undeniable. In this particular context, one who is marked as Black experiences significantly more impediments, put in place by our social and political history, to educational and professional institutions than someone who is marked as White. Affirmative action, on the view I am proposing, is not repayment of a debt owed to a race, rather, it addresses individuals whose experience has been impeded by social and material realities that are the legacy of historical events that we currently consider to be unjust. It is an acknowledgement of the way in which racialization continues in the context of our society, and an attempt to make educational institutions more accessible to those whose membership to racialized groups has negatively impacted their experiences and opportunities.

One might object here, but why should we go out of our way to help those whose experiences are limited by racialization, when there are many contingent factors that might negatively impact one’s experiences or opportunities? I think there are several good answers to this question. First, I would like to point to the fact that the racism that exists in our society today is the historical legacy of a biological essentialist view of race that
we think is incorrect and morally unsound. So, we might think that justice demands that we make a special effort to counterbalance the existing social and material impediments that obstruct the opportunities of those who belong to racialized groups. A second, perhaps less tenable answer, is that the racial groups that exist in our society are not fixed. We might hope that by introducing members of racialized groups (who are assumed to occupy a particular social position) into educational institutions (from which they have historically been excluded) we will be able to disrupt the current patterns of racialization, and eventually our society will be free of these categories.

Third and finally, I would like to draw an analogy between people whose experience is impeded because of the way in which their bodies are marked, and people whose experience is impeded because they are physically handicapped. As a society, we allocate a significant amount of public monies to making buildings accessible to the physically handicapped, even though we are not responsible for the fact that they are disabled. We still think that we should go out of our way to make sure that their experiences are not unnecessarily limited, because we think it is the decent thing to do. I do not want to suggest that belonging to a racialized group is, in and of itself, a handicap, nor that minorities need special assistance because they cannot compete on an intellectual level with those who are racially marked as White. I doubt that suggestion is even intelligible given the definition of a racialized group that we are working with, wherein race does not refer to anything other than gross morphology. Rather, I want to suggest that individuals marked as belonging to racialized groups encounter social impediments, and they are handicapped by a limited experience and limited opportunities because of the current social hierarchy. So we might think that affirmative action is the social equivalent of building a ramp to the educational institutions that have been inaccessible to minorities because of the way in which their experience has been negatively impacted. Even though we might not be personally responsible for the social obstructions, we still believe that assisting one to get around the obstructions and, eventually, helping to demolish the obstructions is at least the decent thing to do, even if it is not required by justice. I’d like to revisit a quote from Scalia, which seems to support helping minorities because they have special needs:

This is not to say that I have no obligation to my fellow citizens who are black. I assuredly do—not because of their race or because of any special debt that my bloodline owes to theirs, but because they have (many of them) special needs, and they are (all of
them) my countrymen and (as I believe) my brothers.

But one might bring up another objection here: When we build ramps to buildings we are not restricting the access of people who do not need the ramps. In other words, affirmative action is different because it helps out minorities at the expense of non-minorities. Is there not still the right of the white male to consider?

As I (by way of Dworkin) argued earlier, I think that the only possible right that the white male could have that might be violated would be a right not to be excluded on the basis of race, *where race implies a biological essence*. On the defense of affirmative action I have proposed, race is understood to be membership to a racialized group, as described by Haslanger, which does not invoke biological essences. He has not been judged on the basis of contempt or prejudice against his race, nor has he been judged as part of a biologically linked group. He has been classified as White because, in response to the way in which his body is marked, he has not experienced the systematic subordination that people who are marked in other ways have. Scalia might object that not all people who are racially marked as white have experienced the same amount of privilege, some recent white immigrants might be more disadvantaged than a well-established, highly educated black family. While it is true that at one time in this country there were Irish, German and Italian racialized groups, that is not the case in our culture today. The problem affirmative action addresses is not one of historical disadvantage, but rather, one of current disadvantage in one’s experience because of the racialized groups that are operative in society now. In other words, racism and the way it negatively impacts people’s experiences at this moment are the specific problems we want to address. Because, in the current context, Irish, Italian and Germans all are marked as belonging to the White racialized group, currently the most systematically privileged group of all, they do not encounter the same kinds of racist obstructions that exist for members of the Black racialized group.

Scalia might counter here that we have an important right not be excluded on the basis of our racialized group. As I said before, what we think is wrong with racist policies such as the Jim Crow laws is that they are premised on a false assumption that there is a meaningful biological distinction between the races. Thus, it seems that judgements based on race are a violation of an important moral right if and only if they are premised on the biological notion of race. Is it possible to discriminate on the basis of racialized groups, such as AA does, in a way that we would say is a violation of rights? I don’t think so, because any attempt to attribute inherent
hierarchy to racialized groups (in such a way that we would say is discrimina-
tory, such as forcing blacks to ride on the back of the bus) is impossible, be-
cause it would have to invoke the biological distinction between the races in order to be intelligible. In other words, one cannot hold contempt for
or discriminate against a person belonging to a different racialized group, be-
cause such an attitude of contempt automatically premises a meaningful
and inherent distinction between the races. Being excluded on the basis of
membership to a racialized group, as Bakke was excluded from Davis Medi-
cal School, isn’t a violation of an important moral right, just as using public
funds to build ramps for handicapped people isn’t a violation of the rights
of able-bodied persons to have their tax dollars benefit them. It is just the
unfortunate side effect of our positive attempts to make public institutions
accessible to everyone.

There is another objection one might raise here, and it is two-fold. First,
on what basis can we determine that every person who is racially marked
has a limited experience? It seems like things have progressed in the past
twenty years, and it is not obvious that, for instance, a Black person from
a wealthy background experiences any social impediment. To this objec-
tion, I can only say that this is an empirical question, but it is not clear how
one could obtain enough definitive evidence to successfully answer it. The
anecdotes and personal accounts I have read written by highly educated,
successful Black people seem to suggest that the negative impact of racism
is experienced, to some extent, by everyone. I quote Appiah,

> What binds the middle-class African-American to his dark-
> skinned fellow citizens downtown is not economic interest
> but racism and the cultural products of resistance to it that
> are shared across (most of) African-American culture.\(^{12}\)

It may be true that racism does not affect every racialized minority
member in this country, but I find that difficult to believe. In any case, should
we abandon affirmative action because it might inadvertently help a few
people who have not encountered racism? Dworkin suggests that the goal
of affirmative action is both to eliminate racism and racial classification in
our society, as well as to assist individuals who are currently impeded by the
racialized social landscape. It seems like, in this case, the price of helping a
few people who have not been victims of racism, in order to work towards
eradicating racism for everyone, is not too high a price to pay.

The second part of this objection asks how we can settle for the five
broad racialized groups, when in fact racism is much more variable. People
are further disadvantaged within the Black racialized group if their skin is especially dark, if their features are less delicate (i.e., less similar to typically white features), or if their vocabulary exhibits less education. That is to say, within the five identified racialized groups in our society, the social and material realities of racism are experienced much more profoundly by some. How do we account for this, and how can affirmative action adequately address this problem?

I do not have an answer to this objection at the moment. I think that within Haslanger’s description of racial identity, there might be much more room to account for this disparity in the experience of people within the same racialized group. This is an idea that needs further development, obviously, and I regret that I cannot treat it adequately in this paper. I doubt it would be possible to establish a system that takes into account the individual level of harm that racism incurs and then compensate for it; we would end up with a reductio not unlike the one Scalia presented. It seems like the defense of affirmative action I have proposed might not work, because I did claim that the main problem it was supposed to address is one of current disadvantage experienced by racialized minorities, and if these are highly variable, I am subject to the objection that affirmative action cannot adequately accomplish the goal it is justified by. I can only say that, addressing racism at the level of racialized group membership is (1) the only practical way to proceed, and (2) that what I attempted to offer is a justification for affirmative action as it is effective now, not as it perhaps should be.

Conclusion

The defense of affirmative action I have given above is far from complete. It seems that once we begin to question the conception of race within the sphere of affirmative action, we are confronted with the ways in which even a socially constructed notion of race cannot adequately capture the phenomenon of racism and racial discrimination. At the very least, I hope to have shown that the question of how we should conceive of race is central to an effective discussion about whether affirmative action is justified, or how it should be instituted. In conclusion, I think that Scalia’s objection to affirmative action, and the general objection that racial classification is never justified, is addressed to a biological notion of race that is not operative in affirmative action. The social reality of race makes a socially constructed notion of racialized groups imperative if we hope to overcome racism, rather than just sweeping it under the rug and denying
its existence, leaving those outside of the ivory tower to deal with it on the level of everyday experience.

NOTES

1 Thomas Bakke sued the University of California, claiming that because the dual track admissions program at the Davis medical school had reserved 16 places for qualified under-represented minorities, it had unjustly discriminated against him. (He is white.) The Supreme Court heard the case in 1977.

2 Dworkin warns that we should not confuse the judgment that race may be socially useful with the judgment that one race is inherently superior to another. In this case, it is not incorrect to make a judgment of merit based on race, he claims, because it is not assigning an essentialist hierarchical value to skin color.


4 Although, the Bakke case itself, and thus Dworkin and Scalia’s papers, may have preceded any worthwhile treatment of the issue, as Haslanger and Appiah’s papers are relatively recent.


Kant, Immanuel On the different races of human beings

6 Appiah, “Illusions of Race”, In My Father’s House, p.35-38

7 Ibid., p.39.

8 Ibid., p.37.


10 Ibid., p. 6-7.

11 Appiah, “African Identities”, In My Father’s House, p.179
ALEXANDRA KHREBTUKOVA

Ontological Privilege: A Defense of Natural Rights

"Let be, be finale of seem."

~Wallace Stevens, "The Emperor of Ice-Cream"
As a basis for the evaluation of a given political system, critics sometimes hold that there exists a system of ‘natural’ or ‘human’ rights which is prior to any legal code and against the backdrop of whose standard the justice of particular political entities is to be assessed. Naturally, the validity of such a standard depends upon the actuality of its pre-political existence. In his *History and Illusion in Politics*, Raymond Geuss argues that the existence of pre-political positive ‘natural’ rights (as opposed to merely moral dispositions) is simply illusory. Aimed at the evaluation of this claim, the present discussion shall proceed as follows: Geuss’s argument against the existence of pre-political ‘natural’ rights is presented and is subsequently countered with an argument for the existence of such a right grounded in the ontological origin of the concept of “right.” A political theory based upon this notion of ‘natural’ right is then outlined.

I.

Political theories that base themselves on the notion of natural rights claim for themselves a transcendental standard “which would allow one to criticize particular existing legal and political systems”(142). In order to determine the truth of this claim, it is imperative to discern whether the concept of a natural right has an ontological foundation or is simply illusory. Thus Geuss concerns himself with the question of a right’s realization—when does the term “right” describe an actual and not just a fictitious entity in the world?

His argument maintains that a right is realized only as long as it is enforced: “I have strongly suggested [that] essential to the existence of a set of ‘rights’ [is] some specifiable and more or less effective mechanism for enforcing them”(143). Thus the right to vote, for example, exists as a reality only as long as there is a real and present system of redress available in the event that the right is thwarted. A natural right, however, is presumed to reside within a realm which transcends a given legal and political system—it is “a point of view rooted in ‘nature’ which was prior to and independent of the vagaries of particular social, legal, and political systems”(142). As such, the concept of a natural right, if it in fact has an ontological foundation in the very constitution of personhood, calls for an enforcing mechanism that lies outside the realm of politics. This conclusion, which sets up the major premise to Geuss’s main argument against any ontological validity to the concept of natural rights, thus follows analytically from his very notion of natural right: in order for something called a “natural right” to truly be a positive right in the ordinary sense of the term, it (a) must be assured of a
mechanism of enforcement, which places certain obligations upon others with respect to the right-holder; and in order for that right to be natural in the present sense of pre-political and pre-legal, (b) the realm in which it is enforced—whose presence constitutes it as a right—must also be pre-legal and pre-political. Hence, in order for a truly natural right to exist, both (a) and (b) must hold of it.

Geuss goes on to argue, however, that the enforcement of any and all rights, including any deemed to be ‘natural,’ is always ultimately political, denying the possibility of the presence of condition (b) above. He points out that “‘Group X has a natural right to Y’ in contemporary political discourse...usually means that they do not have a (legal) right to Y but we think they ought to”(144). Outside the realm of legal enforcement, the notion of a ‘natural right’ is thus a moralizing conception which describes a positive right only to the extent that it is validated by a pre-political enforcing agency. Geuss maintains that without the assumption of “a (Christian style) creator [who] exists and takes care to enforce the rights”(145), all moral claims to ‘natural rights’ deteriorate into mere moral beliefs, lacking any inherent mechanism of enforcement. The realization of such moral beliefs into a system of positive rights depends on the political power of their holder—“[i]f we have enough strength we can make others care about our moral beliefs”(145) to a degree sufficient for the implementation of a legal system for their enforcement. Thus, Geuss argues, any kind of actual enforcement of an attribute claiming the status of right requires a political institution and hence cannot satisfy condition (b) above.

The argument takes the structure of a Modus Tollens:

1) If it is a natural right, it must satisfy certain conditions such that (a) it is enforced by an enforcing mechanism which obligates others in relevant ways with respect to the right-holder, and (b) the enforcing mechanism is pre-political.
2) There is no such thing as a pre-political mechanism of enforcement.
∴ There is no such thing as a natural right (or such set)

Within the post-Christian world view, then, a given ‘right’ describes a real entity in the world only in so far as it is politically constructed: “If we do implement certain ‘rights’ then, of course, they will exist just in the way in which all other positive legal rights exist, but they will exist because we have made them exist’”(145). As such, the notion of a ‘natural’ — that is, pre-
political—right looses all footing. For if all rights “exist because we have made them exist,” then the transcendental standard of political evaluation claimed by a theory of ‘natural rights’ is simply an illusion.

II.

All political systems, however, consist of a number of real persons. The social entity, unless conceptualized in the abstract, is at base a conglomeration of individual people. Before there can be a “we,” there must be two or more “I”s, for the constituents of intersubjectivity are themselves subjective units—an intersubjective plurality is precisely the pluralization of subjectivity. In this way the ontology of the political realm is, in some sense, subordinate to that of subjecthood. As such, all political concepts are necessarily derivative of a prior conceptualization at the level of subjectivity. It is thus imperative to the analysis of a political concept such as that of the ‘right’ to investigate the structure of the concept’s origin at the level of the individual.

Each individual personal entity is initially laden with a bifold structural constitution: a given network of sense data (“world”) and a point of perspective which receives this data (“self”). Further, the “self” must make a series of connections among the sense data in order to give its “world” the coherency required for the “self” to access it and subsequently operate within it. Only when the individual subjective “world” is coherent can subjects begin to operate within it and form political and legal organization.

As pointed out by David Hume¹, however, each and every individual sense datum is unique and distinct from every other. Such being the case, the connections between these sense data are made through the metaphorical, analogizing process of meaning-imposition. Thus, at the most basic level (pre-social), a given set of sensory experiences takes on coherency when, through an abstract realm, a pre-linguistic meaning—such as “unified and continuous object”—is conceptualized to relate the given sense data to each other. As the conceptualization and networking of such meaning-imposition rises in complexity, an individual’s “world” becomes accessible for his/her operation within it.

Simultaneous with this connective networking of its “world,” the point of perspective that is the receptor of the sense data—the “self”—acquires a self-reflective conception of the “I.” For each coherent “world”, precisely because its coherency is not inherent within the sense data constitutive of it but is rather a factor of the meanings imposed by perspective, is necessarily someone’s world. Because the “world” consists of an ever-present flux of
unique and distinct sense data, this *someone* does not properly belong to it (there is no sensory information which singly conveys the presence of *self*) qua “world” and indeed emerges only as the point of synthesis of all the “world”’s individual components—that is, sense data. I say that the point of perspective acquires a *self-reflective* conception of the “I” for the “I,” ontologically grounded in this originary point of perspective, of synthesis, and finds its conceptual ontogenesis as a reflection of the relation in which the originary point of perspective stands to its “world.”

In that any world in which a subject (and hence, thereafter, a polity of subjects) can operate must be *coherent* and a world’s coherency lies in the abstract, meaningful connection of the given sense data, the world, as implied by the political realm, is ontologically identical with its conceptualization. That is to say, independent of the connecting conceptual network imposed upon sense data by the “self,” it makes no sense to speak of a world—the world as we understand it is ultimately that conceptual network. Thus, the relationship which the point of perspective initially bears to its “world,” in being the locus of synthesis and coherency, is such that any given “world,” in so far as it is the “world” that it is, in effect belongs to a point of perspective. The “world” is only a “world” in so far as it is applied to a given perspective, whereupon it is initially synthesized through conceptualization.

From this it follows that, in the process of the accessing of one’s world through representational conceptualization, the point of perspective itself undergoes a conceptualization which reflects its ontological status. The “I” emerges as a concept based on the function of the point of perspective—in its synthesis of individual sense data to make a coherent “world,” the point of perspective in effect appropriates its world, conceiving itself as the subject—the “I.” Thus, the conceptual ontogenesis of the “I” lies precisely in this appropriation of world. In this way, the condition of the possibility of the “I” is that the “I” claims its world: the concept of the “I” is initially found in the conceptualization of “*my* (coherent) world” (in the broadest sense of “*my* general reality”).

This *primordial claim* at the heart of the “I”’s ontology simultaneously constitutes the first and most foundational instance of the concept of “claim.” The interrelations in which it stands to other conceptualizations simultaneous with its own (e.g. that of the “I,” the “world,” etc.) set up certain conceptual criteria which must be respected, to an extent, in the concept’s further development so that the overall structure of representational reality remains coherent. Because of this, the development of the ontologically posterior *political* concept of “claim” and that of the derivative concept of a political “right” (which is something to which the holder has a legitimate *claim*) must both find their conceptual origins in the *primordial claim* within which the
“I” emerges. Thus, the notion of a ‘natural right’ that in actuality describes an ontological property of subjecthood that is “prior to and independent of the vagaries of...political systems”(142) is one that attempts to capture precisely this primordial claim. 

And so, in order to fully understand the aim and constitution of a ‘natural right’ in the political sense, it is helpful to map out the content of the primordial claim of the “I.” What exactly is it a claim to? In that the “I” arises out of the ontological dependency of a coherent world upon perspectival meaning-imposition, such that an activated world is always my (or someone’s) world, the primordial claim inherent in the origin of the “I” is a claim to the construction of one’s own reality. For the connections among sense data that arise as a result of cohesive meaning-imposition (and proceed to structure a world proper) do not do so through absolute necessity. Whether a particular set of sensory impressions is interpreted as food, tool, or sacred object, for example, is not necessarily mandated by the material information provided by the sense data. Nor is it logically demanded that these initial, pre-linguistic, interpretations of a “self” cannot occur, and differ in their content, prior to the imposition of any social norms and conceptions.

The “I”’s primordial claim to the construction of its own reality is thus enforced through the absolute necessity of its truth: for both the “I” and the “reality”—the coherent world—in that they are meaningful impositions upon the raw data of the “world,” exist only in so far as they are originally constructed through individual perspective. Because the social/political realm—as a “we”—is dependent upon the existence of “I”s, any social reality is ontologically rooted in this prior construction of an individual reality. As such, the claim to the construction of one’s own reality is absolute in the sense that, even if such reality-construction is later socially thwarted, the condition of the possibility of the very system of society which aims to impose a social reality is still an initial individual reality-construction.

Once a given subject’s world has in this way been accessed, there does, of course, come a time when individual worlds begin to interact with and shape one another. For, as Geuss is quick to point out, “all human beings are still born as members of groups”(142). Often at this point the subject’s newborn reality rapidly deteriorates into one for which he can no longer take credit, but which is furnished for him by his surrounding society through things like education and culturalization. Nevertheless, the subject exists as such in virtue of the primordial (but perhaps now latent) claim to the construction of one’s own reality. In this way a claim is ontologically enforced (in the sense that it cannot be otherwise), while lacking actual political enforcement.
III.

If the primordial claim to the construction of one’s own reality is dubbed the individual ‘natural’ right of every subject, then a theory of political evaluation based upon the notion of a ‘natural human right’ simply points to the ultimate contingency of all political realities upon the individual realities of its subjects.

The primordial claim out of which the political concept of “right” derives is not altogether content-less—for it retains the general content of a right to the construction of one’s reality—but it does not prescribe a *particular* reality for the subject. Because of this, it is misleading to give, as Geuss does, the ‘natural’ right a concrete *specific* form, such as “a property right” (140) or “a right to a minimal level of welfare” (141), etc. In doing so, we actually negate the natural right inherent in the primordial claim—the right to the construction of one’s own reality. To specify the content of a natural right *denies* the subject her own construction thereof.

Instead, it is better to formulate the ‘natural’ right of a human being as simply the ultimate say in the conceptualization and operation of one’s own world. As such, the ‘natural’ right may not always (or ever) be politically enforced, and yet the ever-present instability and impermanence of all political systems continuously attest to its existence (and ontological enforcement). This ontological enforcement denies premise (2) of Geuss’s larger argument against any ontological grounds for the existence of natural rights (see Section I above)—that there do not exist any mechanisms of pre-political enforcement.

Geuss’s argument for this premise invokes a situation in which the subject (or group thereof) has no *legal* right to something but claims it on the grounds of possessing a *natural* right. He asserts that this “natural right” cannot be enforced—and so is not a positive right proper—until it is properly inserted into the legal system, whereby vindication for its denial will be assured. But if the natural right is the right to the construction of one’s own reality, and if the reality fabricated by the present legal/political system doesn’t suit the subject, then the right is enforced as soon as it is exercised—in the disavowal of the present legal/political system. Take the infamous case of Henry David Thoreau, who refused to pay taxes at a time when he knew the money would be used for a cause he did not wish to support. *Legally*, he had no right to do this—he did not have a (legal) right not to pay his taxes. Yet he had an *ontological* right to fabricate his own reality, which was ontologically enforced the moment that it was exercised, in the rejection of the reality pre-fabricated for him. He had no legal right not to pay his taxes and yet he didn’t pay his taxes and did not support the cause
he fought against.

Thoreau was, of course, imprisoned for such behavior, and here one might well ask how a subject in Thoreau’s position might exercise her ontological right to the construction of her own reality if she wished to construct one which involves both not paying taxes and not being in jail. The issue at stake points to the fact that the enforcement of a right involves a reciprocal obligation on the part of others to respect that right. My legal right to free speech, for example, implies your legal obligation to respect that right, except in certain legally specified contexts.

But here it must be remembered that the legal system is itself a constructed reality, fabricated by a number of subjects as their communal “world.” And as Geuss himself points out, which “rights” make it into explicit legal configuration depends upon the political power of their human advocates—“[i]f we have enough strength we can make others care about our moral beliefs” (145) enough to make them legally obligatory. The creation of the legal system, within which certain paths of rights and obligations are momentarily (for change is always an immanent possibility) rigidified, is a coming together of “worlds” to make a communal “world.” As such, in order to gain precedence over others (if that is the desired effect), the holder of any given “world-view” must possess a certain amount of power of influence, in one way or another, by way of which it comes to be accepted.

This same principle operates outside of (and prior to) the legal or political sphere. For, as previously mentioned, individual worlds interact with and shape one another. Stemming from the ontogenesis of subjecthood, each has within it a latent right to and capacity for self-construction. Who shapes whom and how is precisely the drama of the political arena, in which “worlds” attempt to come together in a common conceptualization. Since there does not exist an absolute objective standpoint from which to normatively judge one “world” over another, the obligation of others to respect a given subject’s right to self-construction lies in the extent to which a world-view is successful in influencing others toward its acceptance. This does not mean that the subject’s right to the construction of her own reality is not enforced if her “world” is not accepted—it is enforced as soon as she rejects the world provided for her. It does mean, however, that the exercise of this right in contexts involving the active participation of other subjects requires a certain amount of either common ground among the individual subjects or power of influence.

Nevertheless, this supra-legal right is frequently exercised, especially in political contexts where the subject has no legal say in the construction of his reality. The underlying tone of all political change and political revolt is precisely the activation of one’s primordial claim to the construction of
reality. Any rejection of the prescribed reality presupposes the individual’s right to choose his world-view, a right that is enforced through the very act of that rejection.

And hence it seems that, contrary to Geuss’s proposal, there does exist a ‘natural’ right (in the sense outlined above) that is inherent in the very core of subjectivity, though not in the content-oriented sense that Geuss attributes to it. As such, it is, indeed, at the very heart of the constructive mechanisms of politics.

NOTES

1 Hume, David: Treatise on Human Nature, 1737; the Appendix.

WORKS CITED

No, take more:
What may be sworn by, both divine and human,
Seal what I end withal! This double worship,
Where one part does disdain with cause, the other
Insult without all reason, where gentry, title, wisdom,
Cannot conclude but by the yea and no
Of general ignorance, – it must omit
Real necessities, and give way the while
To unstable slightness: purpose so barr’d,
it follows,
Nothing is done to purpose. Therefore, beseech you, –
You that will be less fearful than discreet,
That love the fundamental part of state
More than you doubt the change on’t, that prefer
A noble life before a long, and wish
To jump a body with a dangerous physic
That’s sure of death without it, at once pluck out
The multitudinous tongue; let them not lick
The sweet which is their poison: your dishonour
Mangles true judgment and bereaves the state
Of that integrity which should become’
Not having the power to do the good it would,
For the in which doth control’.

~ William Shakespeare, Coriolanus, Act III, Scene I
NOAM JACOB

Nozick, Cohen, and Scheffler: Entitlement, Ownership, and a Meaningful Life

With Responses By:
ADAM SASSOON
DAVID JONAS COHN
DAVE SHUKLA
Locke’s conception of the “state of nature” affords individuals the freedom to arrange their possessions, persons, and priorities however they see fit provided they act in accordance with the law of nature: “no one ought to harm another in his life, health, liberty, or possessions.” However, within the state of nature certain “inconveniences” will inevitably arise; individuals may transgress the law of nature causing the injured parties to seek retribution and, as a result of personal enforcement of one’s rights, may eventually lead to an endless series of retaliations and exactions of re-compensations. Locke seeks to remedy these foreseeable inconveniences by stipulating the need for some form of a civil government. Exactly what arrangements need to be made within the state of nature—i.e., what form of governing institution should be established in order to avoid, decrease, or reduce the severity of these inconveniences is precisely the aim of Robert Nozick’s investigation in *Anarchy, State, and Utopia*. While Nozick attempts to overcome the difficulties resulting from other theories of justice, his “entitlement conception” yields problems of its own. This paper will focus on presenting Nozick’s theory and the arguments he provides for it, along with discussing the objections of G. A. Cohen and Samuel Scheffler, particularly how the notion of a meaningful life plays a role in determining the form of individual rights.

I. Robert Nozick’s Entitlement Conception of Justice

In considering various forms of civil government, Nozick argues that a minimal, or “night-watchman” state, limited to protecting all its citizens against violence, theft, coercion, and to enforcing contracts, etc., is the most justifiable state. Nozick initially distinguishes the minimal state from what he calls the ultraminimal state: a state that provides protection and enforcement services only to those who purchase its protection and enforcement policies. The proponent of the ultraminimal state, Nozick notes, confronts an inconsistency while asserting that the sole legitimate function of the state is to protect against the violation of its citizens’ rights: while the protection and non-violation of rights takes such precedence, the ultraminimal state does not provide this protection to everyone. In other words, how can the proponent of the ultraminimal state justify the lack of protection of only some peoples’ rights if the matter of rights-protection is of such importance? Like many prominent theories of justice, as Nozick explains, this question assumes that a moral concern must take the form of an end state or goal that results from certain courses of action. Utilitarianism, for example, does not
properly take into account rights and their nonviolation; it instead, renders them a derivative status. Accordingly, the goal of the theory is to maximize utility, in this case perhaps to maximize the amount of protection of rights (namely, by minimizing the violation of rights). Postulating some overarching goal or end state poses significant difficulties, however. The classic example of violating an innocent person’s rights in order to minimize the amount of total rights violated presents a serious question to proponents of an end-state theory whose goal is to minimize the amount of violated rights. Nozick argues that if rights are viewed as side-constraints, constraints upon performing certain actions, rather than being incorporated into the end state of some goal, these questions can be avoided.²

Nozick’s justification for this claim, while limited, along with various considerations of the inadequacies of other theories, leads him to adopt his minimal state view accompanied by the “entitlement conception” of justice. Nozick contends that individuals are ends in themselves; they cannot be used as means to achieve any final goal without their consent: “individuals are inviolable.”³ Thus, individuals are the primary entities of the state; no overall social good exists as a justifiable goal, for which the rights of some may be violated to ensure its preservation.

After briefly considering the status of rights and different conceptions of justice (utilitarianism, the ultraminimal and minimal state), Nozick proposes his “entitlement conception” of justice, and then discusses alternative views that he claims are unjust. As stated earlier, Nozick contends that any state more extensive than the minimal state violates people’s rights. The principles by which the state abides in order to maintain its minimalist status are outlined in the entitlement conception. The theory consists of three principles roughly derived from the Lockean proviso⁴: 1) a principle of justice in acquisition, 2) a principle of justice in transfer, and 3) a principle for the rectification of injustice. The principle of justice in acquisition addresses the issues of how an individual becomes the owner of an object previously un-owned. An individual may come to own some object that is not the property of some other individual if a) the individual “mixed” his labor with the object and b) no one’s situation worsens as a result of this individual’s owning the object. The principle of justice in transfer concerns the transfer of property from one person to another. An individual is entitled to an object that was the legitimate property of some other individual if the other individual voluntarily transferred the object as a gift or in exchange for some other object. The principle of rectification of injustice uses historical information about past situations, and injustice as defined by the first two principles, and attempts to rectify these injustices by “making use of its best
estimate of subjunctive information about what would have occurred (or a probability distribution over what might have occurred, using the expected value) if the injustice had not taken place” (Nozick, 152-3). While Nozick omits the specific details for each of these principles, the general theory of justice in holdings stipulates that a distribution of holdings is just if everyone is entitled to their holdings according to the three principles of justice.³

Nozick contrasts his view from other theories that he believes fail to be completely just. He categorizes these other views according to two distinctions. The first distinction concerns his earlier conjecture about end-state principles of justice and historical principles of justice. The second distinction is made within each of the first two distinctions, and concerns patterned vs. unpattereded distributions of holdings. End-state or time-slice principles contend that a distribution is determined by some structural principles of justice. A utilitarian who judges between two distributions by determining which one has the greater sum of utility, applies a fixed principle to choose the more “equal” distribution. On this view, when judging the justice of a distribution, all that needs to be considered is who ends up with what. Note, however, that any two structurally identical distributions are equally just: A having ten and B having five, or A having five and B having ten are structurally identical distributions since “they present the same profile, but perhaps have different persons occupying the particular slots”(Nozick, 154). In contrast, Nozick’s entitlement conception of justice in distribution is historical; whether a distribution is just depends on how it came about. A distribution is patterned if “it specifies that a distribution is to vary along with some natural dimension, weighted sum of natural dimensions, or lexicographic ordering of natural dimensions” (Nozick, 156). Patterned distribution, therefore, could be defined by the statement: “to each according to his _____,” where one could substitute a variety of different principles, ranging from historical to end-state factors. Unpatterned distributions, in contrast, do not specify such a dimension by which goods are to be distributed. Thus four categories emerge: 1) end-state patterned, e.g. distribute holdings according to I.Q. or “usefulness to society,” 2) end-state unpatterened, e.g. distribute holdings such that a final structure or end state is achieved, but not necessarily according to some underlying pattern (utilitarianism falls into this category), 3) historical patterned, e.g. distribute holdings based on some historical pattern such as moral merit, and 4) historical unpatterened, Nozick’s entitlement conception fits into this category.

Nozick explains that end-state theories of justice fall prey to the difficulties that stem from not considering individuals as ends in themselves, as mentioned earlier, but category 1) has the additional problem of lacking
justification. Regarding punishments and penalties the notion of desert is employed. Nozick argues that regarding appropriations we should think it relevant to question whether an individual deserved to receive the share he was allotted. Individuals born with certain natural talents cannot be said to have deserved them. Therefore, rewarding such individuals with a share of a distribution that is determined by patterns of natural ability or usefulness seems arbitrary.

Nevertheless, patterned distributions, whether end-state oriented or historically motivated, play a prominent role in theories of justice. Nozick believes he can overcome the issues of patterns (and those arising from end-state theories of justice) by demonstrating that these distributions infringe upon individual rights. His Wilt Chamberlain example illustrates a change in a just distribution $D_1$ to another distribution $D_2$ through the means of a voluntary transfer of money from a certain group of people to one person. Suppose that basketball teams greatly demand Wilt Chamberlain, and in recognition of this he agrees to sign a contract specifying that he receive twenty-five cents from the price of each admission ticket. At a given season, one million persons attend his home games, and Wilt Chamberlain earns $250,000, a much larger sum than the average income and larger even than anyone else has. “Is he entitled to this income? Is this new distribution $D_2$, unjust?” (Nozick, 161). Under Nozick’s entitlement conception such a transaction is completely just (provided the principles of justice under the entitlement conception are followed) and the newly resulting distribution is likewise just. Under a patterned (or other end-state) distributive theory of justice such a situation would be prohibited since the voluntary transfer upsets the pattern, end-state distribution, or whatever the final goal of the theory. With this example Nozick argues: “no end-state principle or distributional patterned principle of justice can be continuously realized without continuous interference with people’s lives” (Nozick, 163).

It is important to emphasize that Nozick’s theory is primarily concerned with the way in which holdings are acquired and transferred, similar to a complete capitalist or free market system. One should not make the mistake of assuming that the goal of Nozick’s theory is to ensure a just distribution for all parties involved. This mistake classifies Nozick’s theory as one that aims at maintaining some end-state, or a moral goal—a just distribution. Nozick claims that the final distribution will be just if the manner in which it came about is just, but no end-state exists as the object of the theory, rather only procedural principles are needed. If one places priority on the means being just, the end-result will also be just. But if one places sole priority on the end being just, then, perhaps the means will be “justified” (i.e., the
violation of rights will be deemed acceptable because it yields the desired end) in order to ensure the desired end state—an unacceptable situation that Nozick wants to avoid.

II. **G. A. Cohen on Private Appropriation and Joint Ownership**

A significant criticism of Nozick’s entitlement conception derives from G. A. Cohen’s\(^7\) argument against Nozick’s principle of justice in acquisition, focusing on its second clause. Cohen insists that any appropriation will make someone *worse* off, for no other reason than that someone will no longer be able to appropriate the now-appropriated item: “It is clear beyond a doubt that an appropriation of private property can contradict an individual’s will just as much as levying a tax on him can” (Cohen, 90). If the violation of one’s will is a measure for a theory that is supposed to be based on liberty, then no private appropriation could meet the conditions of a suitably formulated Nozickian proviso, since even if a latecomer who finds no unappropriated resources left were to be compensated by some form of material wealth, this compensation could not amend the violation of the latecomer’s will. As Cohen argues: “Nozick disallows objectively paternalist use of people’s private property. But he permits objectively paternalist treatment of people in other ways. For, since he permits appropriations that satisfy nothing but his proviso, he allows A to appropriate against B’s will when B benefits as a result, or rather as long as B does not lose” (Cohen, 89). In this respect, even if the claim above is not accepted, Cohen observes that private appropriation may well not *worsen* (from Nozick’s perspective) the situation of anyone, whom the appropriation excludes from being able to freely use what was appropriated, when Nozick’s common ownership is considered the “base-line” for comparison.\(^8\) But Cohen argues, however, that it is arbitrary to restrict the base-line to the case of common ownership when trying to determine whether the situation of others has been worsened by private appropriation, since there are other “intuitively relevant counterfactuals” to common ownership that could be considered the base-line, and where the situation of those whom private appropriation excludes does worsen if their situation in common ownership is compared to these new possibilities.

Cohen explains his point with a simple two-person world inhabited by A and B, in which there is common ownership of its finite quantity of land. Suppose A and B each subsist by obtaining individual means of nourishment “without obstructing the sustenance-drawing activity of the other” (Cohen, 79), in a case where neither has privately appropriated the land. In this situ-
ation, imagine that \(A\) obtains \(m\) from the land and \(B\) obtains \(n\), where \(m\) and \(n\) are bushels of wheat or gallons of moose milk from wild moose which neither \(A\) nor \(B\) owns. Now suppose that \(A\) appropriates all the land— or an amount that leaves \(B\) with less than enough to survive. As compensation, \(A\) then offers \(B\) a salary of \(n + p\) (where \(p > 0\)) to work on the land as his employee and \(B\) accepts. Now suppose that \(A\) obtains \(m + q\) from the new agreement (where \(q > p\)) due to, according to Cohen, “the productivity of a division of labour designed by \(A\) who is a good organiser”\(\text{(Cohen, 79).}\)

Cohen terms this situation the “actual situation.”

Nozick considers common ownership, wherein \(B\) obtains \(n\) in the above scenario, as the appropriate base-line with which to compare the “actual situation” in which \(B\) obtains \(n + p\). Thus if common ownership is made the base-line for comparison, then it seems as if \(A\)’s appropriation of the land has not actually worsened \(B\)’s situation. As mentioned above, however, Cohen claims that there are justifiable different forms of ownership that can be considered the base-line with which \(B\) can compare his post-appropriation situation and determine that his situation worsens in the actual situation when compared to these other possibilities: 1) a situation in which \(B\) appropriates the land instead of \(A\), and 2) a situation in which the land is *jointly owned* by both \(A\) and \(B\), and as a result, any decisions made regarding the land must be consensual. In both these cases, Cohen argues, it is possible that \(B\)’s situation will be better than in the actual situation. Consequently, even if \(A\)’s private appropriation of the land can improve \(B\)’s situation when compared to common ownership, this same appropriation can be said to have worsened \(B\)’s situation when it is compared to the two cases above. Nozick’s selection of common ownership as a base-line is, therefore, arbitrary.

Cohen examines three possibilities regarding \(B\)’s organizational talent to explain his argument. Consider the first situation of the two mentioned above where \(B\) appropriates the land instead of \(A\). \(B\)’s organizational talent is either (i) the same as, (ii) greater than, or (iii) less than \(A\)’s talent:

i) If \(B\)’s talent is the same as \(A\)’s and \(B\) appropriates the land instead of \(A\), then by invoking the same guidelines established by the scenario in the actual situation, \(B\) (rather than \(A\)) would be able to introduce the new division of labor, persuade \(A\) to work for a salary \(m + p\), and \(B\) would have obtained \(n + q\). In this situation, as a result of appropriating the land, \(B\) receives more than he receives in the actual situation, since recall that in the actual situation, with \(A\)’s appropriation of the land, \(B\) receives \(n + p\), whereas now \(B\) receives \(n + q\) (where \(q > p > 0\)). The actual situation did not violate Nozick’s proviso, but when the actual situation is compared to this new arrangement \(B\)’s situation is clearly better in this situation than in the actual situation.
We can therefore say that when comparing the actual situation to this new arrangement, $B$’s situation worsens if the actual situation obtains. “It should now be clear that Nozick’s proviso is too weak” (Cohen, 80).

ii) If $B$’s organizational talent is greater than $A$’s and $B$ appropriates the land instead of $A$, then both $A$’s and $B$’s situation is better than in the actual situation. “Nozick’s proviso is nevertheless, satisfied, since whether or not it is satisfied is unaffected by anything that might have happened had $B$ appropriated. And this means that Nozick’s condition licenses, and protects, appropriations whose upshots make each person worse off than he need be” (Cohen, 81).

iii) If $B$’s organizational talent is less than $A$’s and had $B$ appropriated the land, he could not have directed $A$ so as to generate any increase over what gets produced under common ownership. Even in this case, Cohen argues that $A$’s appropriation of the land may be said to have worsened $B$’s situation, despite $B$’s situation having been improved had he been in a case of common ownership. Cohen’s rationalizes that restricting the base-line to the case of common ownership takes for granted another possibility; namely, that the land before $A$’s appropriation was owned by no one, as in common ownership, rather than jointly owned. When land is owned in common, Cohen explains, each can use it however he chooses, provided that he does not interfere with its use by others; under common ownership of the land no one owns any of it. In contrast, joint ownership of the land prevents one’s use of the land without the consent of both. Joint ownership of the land provides $B$ the “right to forbid a to appropriate even if $B$ would benefit by what he thereby forbids” (Cohen, 84). Consequently, $B$’s right to prevent $A$’s appropriation of the land can subsequently allow $B$ to extract from $A$ a greater share of the increased output (in the actual situation the sum of $q + p$), generated by the new division of labor introduced by $A$, than $B$ would have received from $A$, had $A$ been allowed to appropriate the land without $B$’s consent under common ownership. Thus, even where $B$’s talent is less than $A$’s talent, his situation can be worsened by $A$’s private appropriation of the land and an introduction of a division of labor that yields $B$ more than he would obtain under common ownership, since $B$ would have been able to obtain still more from the introduction of a division of labor had the initial situation been that of joint ownership.

These cases demonstrate that “there will always be some who would have been better off under an alternative dispensation which it would be arbitrary to exclude from consideration” (Cohen, 87). Nozick’s selection of common ownership as a base-line is therefore no better than any of the variations of the actual situation or joint ownership, and as Cohen argues,
joint ownership is justifiably better since it does not allow for massive inequalities, not of opportunity, but of condition.

Various objections to Cohen are available for Nozick as a source of recourse, ranging from Cohen’s misreading of Nozick to his lack of justification for his claims above, to inconsistencies in his reasoning. This objection to Nozick assumes that choosing an appropriate baseline is an integral part of the entitlement conception. However, the notion of entitlements, or rights, and considering the separateness of individuals forms the basis for the entitlement conception. Cohen continues his sustained critique of Nozick along the lines of self-ownership: the view that an individual is entitled to the fruits of his labor in the same way that an individual has rights over the ownership of his own person. Instead of presenting Cohen’s argument against self-ownership, I will focus on a criticism of Nozick that I feel cuts deeper into the heart of the entitlement conception; namely, an attack that identifies an inconsistency in Nozick, which results from the underlying justification that founds the entire entitlement conception of justice.

III. Samuel Scheffler’s Alternative Conception of Rights

As may already be evident, Nozick’s arguments are determined mainly by his consideration of rights as side-constraints. Notice that the Wilt Chamberlain example only poses problems for the other theories of distributive justice if one views rights in this manner, or if one finds any infringement of individual liberty troubling. While Nozick claims that individuals are ends (and should not be “used” without their consent) and that there is no “social good” that takes priority, his justification for such claims is limited, at best. When answering the question of why individuals’ rights cannot be violated for any reason Nozick initially states: “Individuals are inviolable.” This answer merely restates the question; no sufficient reason is actually provided. Only later does Nozick justify this claim. This justification, however, under closer scrutiny is subject to serious criticism.

Nozick explains that the basis for the non-violation of rights and for treating individuals as ends stems from “that elusive and difficult notion: the meaning of life”—the assumption being: only an individual “with the capacity to so shape his life can have or strive for a meaningful life” (Nozick, 50). Presumably, a violation of one’s rights prevents one from living a meaningful life. This assumption, however, while still insufficient and unelaborated (as Nozick admits the need for its clarification on another
occasion), raises another difficulty: why should rights only take the form of constraints? One could perhaps argue that every individual has a right to a share in the distribution of goods that is sufficient enough for that individual to live a meaningful life. If living a meaningful life forms the basis for rights, then within any theory that presumes to protect those rights should be included a method for preserving the conditions of possibility of those rights; namely, a method that ensures that an individual will actually be able to lead a meaningful life. This is the line of argument that Samuel Scheffler\textsuperscript{9} presents with an “alternative conception of rights” by claiming that the rationale that Nozick provides for his view of rights better supports this “alternative view” than it does the entitlement conception. On this view, the function of the state would not only be to protect the rights of its citizens against violation, but also ensure that all its citizens will be able to lead a fulfilling life by ensuring that individuals will have their necessary share of the distribution:

Every person has a natural right to a sufficient share of every distributable good, whose enjoyment is a necessary condition of the person’s having a reasonable chance of living a decent and fulfilling life, subject only to the following qualification. No person has a natural right to any good which can only be obtained by preventing someone else from having a decent and fulfilling life (Scheffler, 153).

In contrast with Nozick’s theory, which provides individuals such stringent property rights, Scheffler proposes that the needy have a right to holdings that are required for living a decent and fulfilling life, and that these holdings cannot be withheld by someone who does not need them for the same capacity. Scheffler contends that Nozick’s reasoning for his own theory of rights is not only consistent with this alternative view, but supports Scheffler’s alternative better than it does the entitlement conception and Nozick’s consideration of rights as side-constraints. As explained earlier, if beings with the capacity to live a meaningful life have rights in virtue of having that capacity, “then presumably the function of the rights is to safeguard the ability of beings with this valuable capacity to develop it” (Scheffler, 158-9). Thus principles that only regulate acquisition and transfer are insufficient for preserving the capacity of individuals to lead a life that is meaningful. According to Scheffler, the alternative conception of rights “alone insures that all the necessary material conditions for having a reasonable chance of living a meaningful life will be met” (Scheffler, 159).
To summarize: Nozick’s “entitlement conception” of justice seems to overcome the difficulties presented by patterned distribution theories of justice as well as end-state theories of justice. Nozick’s principles, however, are problematized by difficulties arising from the underlying assumptions of his argument: 1) the argument for considering rights as side-constraints based on the notion that individuals are “inviolable” is limited; 2) even if one accepts Nozick’s final claim that side-constraints are based on the capacity for each individual to live a meaningful life, Scheffler’s criticism arises, namely, principles that only provide negative rights (i.e. protect an individual’s rights from being violated), and regulate transfer and acquisition are insufficient to guarantee that each individual will be able to fulfill his capacity for living a meaningful life.

While Scheffler’s alternative is presented here in a favorable light, it still requires further examination. In what follows, I present an objection to Scheffler’s view that can take various forms, ultimately relying on the assessment of the basis for rights. After outlining the objection I examine it further by providing two points of clarification that question Nozick’s justification for his conception of rights, and a third point noting a deficiency in Nozick’s argument. Despite the objection to Scheffler and the points questioning it, a possible re-reading of his claim will show that Scheffler’s alternative does not necessarily entail the assumptions on which the objection rests.

An immediate objection to Scheffler’s alternative contends that in order to provide all individuals with the means necessary for them to live a meaningful life a system of taxation would be necessary. Initially, this claim may seem innocuous, but recall that Nozick equates taxation with forced labor (see, note 4), and when presented with this equation, Scheffler’s alternative seems difficult to justify. The equation of taxation with labor has been subjected to serious scrutiny, however, and when confronted by points of disanalogy, the equation may not hold up. These points are as follows. An initial distinction exists between taxation without one’s consent and what one would normally consider forced labor, e.g. chain gangs, slavery, etc. Regarding taxation, the person who is required to contribute a percentage of his salary, presumably to those for whom it has far greater marginal utility, is still free to work as long as he likes and, to a certain extent, at the job he chooses. Moreover, one can argue that differences in earnings generally do not reflect an amount of hours worked or effort spent at a particular job. These differences are more often a function of the initial undeserved advantages that certain individuals randomly have as a “starting-point” in life, e.g., being born into a wealthy family, with a high IQ, or endowed with certain talents. For these reasons one cannot strictly equate forcing an individual to
work extra hours, or expend more effort to make up a difference in salary, with redistributive taxation, which takes a portion of the earnings that are a result of initial arbitrary advantages to provide for the basic needs of those who similarly undeservingly lack those advantages.\textsuperscript{10}

While perhaps instructive, these disanalogies are not sufficient for rejecting the equation of taxation with forced labor, since the source for the equation ultimately derives from the claim that the well-off are coerced to make sacrifices (in this case provide money) for others, sacrifices that do not benefit themselves, and therefore, the well-off are used as means for the benefit of the needy. To understand why this claim would still damage Scheffler’s alternative one must further examine Nozick’s conception of rights and their origin. This objection\textsuperscript{11} to Scheffler’s alternative contends that while Nozick does ultimately base constraints on the capacity of humans to live a meaningful life in the section titled ‘What are Constraints Based Upon?’ he also has a section titled ‘Why Side Constraints?’, in which he argues that the reason for side constraints derives from Kant’s second categorical imperative: “individuals are ends and not merely means; they may not be sacrificed or used for the achieving of other ends without their consent” (Nozick, 31). Along this interpretation of Nozick, he is understood as providing the following two qualifications:

(1) A being is an end in itself, and may therefore not be used as a means iff that being has a capacity to live a meaningful life.\textsuperscript{12}

(2) If a being is an end in itself, then its rights are such that it is never to be used as a means only.

Nozick thus first maintains that individuals should be treated as ends because of their capacity to live a meaningful life and, consequently, the form of rights that individuals possess exist as side constraints in order to protect their status as ends, who may never be treated as means only. If Nozick is interpreted in this manner, then Scheffler’s alternative is not better supported than the rights Nozick recognizes by the justification Nozick provides for them. Furthermore, with this clarification of Nozick, the objection to Scheffler does not necessarily have to take just the form of taxation (the point being that redistributive taxation uses those who are taxed as a means for the welfare of the needy), but rather, recognizes that Scheffler’s alternative may result in the treatment of some as a means for the benefit of others, whether through taxation or obligations to spend time on charitable work. Consequently, this objection claims that while Scheffler's alternative provides
positive rights for the needy in order for them to fulfill their capacity for a meaningful life, his view condones the violation of the rights of others by treating them as means.

Two points of clarification should be made, however, regarding this objection which ultimately determines that the difference between the injunction against treating people as means and the requirement to help others in need is so great that enforcing the first follows a fundamental principle of morals whereas enforcing the second violates such a principle. First, it is not certain, and moreover Nozick has not completely established, that individuals can never be used as means under any circumstance. It has been argued that there is a class of ends, the nondistributive goods, to which proper treatment of persons is an indispensable means. This view claims that we have reasons from the moral point of view to use individuals as means, and in so doing endorse the following propositions: (i) an agent ought to avoid injury to persons, (ii) some ends are such that in order to promote them, it is (causally) necessary to avoid injuring persons, and (iii) an agent ought to promote these ends. For brevity, I will simply limit my discussion to mentioning this view and will not pursue its argument, as it is not the aim of this conjecture (for more on this view, see note 12).

A second point re-examines Nozick’s Kantian basis for rights. As explained earlier, Nozick views his conception of rights as Kantian in character: “side constraints upon action reflect the underlying Kantian principle that individuals are ends and not merely means” (Nozick, 30). This principle derives directly from Kant’s second formulation of the categorical imperative: “Act in such a way that you always treat humanity, whether your own person or in the person of any other, never simply as a means, but always at the same time as an end.” Kant, however, also provided the first formulation of the categorical imperative, which he considered equivalent to the second: “Act as if the maxim of your action were to become through your will a universal law of nature” (Paton, 84). To explain this formulation of the categorical imperative, Kant provides an example of an individual who asks why he should not focus on making himself happy and refuse to come to the aid of others in distress. Kant explains that this person could not will that everyone ignore the cries of others in need, since, perhaps, on another occasion this well-endowed person may be in need of help from others. Kant suggests that one cannot will that some way of acting (such as denying aid to those in need) becomes universal practice if there is a possibility that on another occasion this individual would will the contrary to be the case (Paton, 86). Thus, from this application of the first formulation follows the moral duty to help others in need. As explained above, Kant provided two such
principles that are concomitant to one another. Thus the adoption of one of the two moral principles should require the adoption of the other, given the context and intention of Kant’s project. Nozick’s conception of rights, and the objection to Scheffler that only focuses on the second formulation of the categorical imperative, fail to recognize the first formulation and its association with the second formulation. Consequently, Nozick and proponents of his view would be inconsistent were they to claim that a forced re-distribution of goods violates the Kantian conception of rights since, while making this claim, they would fail to acknowledge that the denial of such a method for helping those in need may violate the very same conception of rights that was in such desperate need of preservation in the first place.

It is important to note, however, that Nozick makes a distinction between moral and political philosophy (Nozick, 6 and 32) and claims that the two realms are not coextensive. What one ought to do is not, for him, the same as what one should be enforced to do. Morality covers a broader realm than enforceable obligations, which Nozick recasts as rights. Correlatively, political philosophy, in contrast to moral philosophy, is concerned with rights and enforceable obligations. For Nozick, in situations where moral considerations conflict with negative rights to life, liberty and property, these latter trump over the moral considerations. In adopting a Kantian basis for rights, however, Nozick flouts this distinction and utilizes a Kantian moral principle as a basis for his conception of rights—but this does not mean that Nozick must accept both the first and second formulations of Kant’s categorical imperative. Rather, Nozick only explains his view in Kantian terms and is not required to accept the reasoning behind their inception, nor does he have to accept their corresponding entailments.

With this in mind, it is not clear exactly why the capacity for a meaningful life must yield that an individual can never be used as a means. Nozick later claims that treating an individual as a means fails to respect his separateness from others, as well as interfering with his ability to lead a meaningful life (Nozick, 33). With this much said, the objection to Scheffler seems to distinguish between what may be two competing principles. On the one hand, Nozick argues that to use a person as a means violates his rights since it does not sufficiently respect and consider that he is a separate individual. On the other hand, Scheffler would argue that failing to provide for a person in need (when one can do so without sacrificing his own ability to live a meaningful life) equally fails to consider and respect a human being in need. In other words, what ultimately demands more consideration: no individual being treated as a means only, or every individual being given a chance to fulfill his capacity for a meaningful life?
Leaving aside for the moment a discussion about this question and the form that rights should take, it may be worthwhile to first consider that Scheffler can perhaps support his own view by noting a deficiency in Nozick’s view that Scheffler’s alternative conception can overcome. Recall Nozick’s initial criticism of time-slice or end-state principles and the disadvantage they have in comparison to historical principles:

We think it relevant to ask whether someone did something so that he deserved to be punished, deserved to have a lower share . . . In contrast to end-result principles of justice, historical principles of justice hold that past circumstances or actions of people can create differential entitlements or differential deserts to things (Nozick, 154-155).

Thus, to borrow a term from Rawls, Nozick considers time-slice or end-state principles problematic since they are morally arbitrary in that they ignore the notion of desert when determining distribution. Individuals that may have an initial advantage due to an arbitrary lot in life such as being born into a wealthy family or endowed with certain talents are rewarded twice by a distribution that only specifies a certain structure or current state of affairs. Later, perhaps to avoid an inconsistency, Nozick refrains from pursuing the notion of desert any further, and instead introduces his conception of entitlement, which he believes overcomes the difficulties raised by desert. Were Nozick to rely on desert to a greater extent in his political conception, the inconsistency would result as follows: Nozick allows parents to bequeath large sums of money to their children without any limitations; an individual who is born to wealthy parents does not deserve that position. Similarly for those who obtain certain employment opportunities due to similar undeserved initial starting points in life, or natural skills with which they were endowed, or for an individual who is disabled—in all of these cases the individual does not deserve his initial position in life. Yet Nozick finds no problem in protecting the arbitrarily elevated status of the individual who receives the favorable lot in life, while providing no aid to those who likewise are arbitrarily deprived. In this respect, Nozick rewards those who have been already rewarded and ignores the comments he made earlier regarding desert.

By relying on the notion of entitlement rather than desert, Nozick can avoid rewarding individuals on morally arbitrary grounds and the inconsistency with the comment made about desert above. In contrast to desert, which cannot, strictly speaking, be attributed to an individual’s talents, posi-
tion in society, etc., and in order to argue against Rawls, who proposes the difference principle to remedy the problem about natural assets and their moral arbitrariness, Nozick argues that entitlement is the only reasonable alternative since it applies to the aforementioned natural assets—the claim assuming that since an individual owns himself, he is entitled to or owns his talents (in contrast to him deserving his talents, or alternatively, as Rawls claims, everyone being entitled to his talents). By substituting entitlement for desert Nozick makes little progress, however, since although the problems arising from desert are avoided, the claim that an individual is entitled to his talents is not firmly grounded. Whereas most would agree that an individual owns himself, it does not necessarily follow that he is entitled to his talents, position in society, or other natural assets with which he may be endowed. This claim is even more problematic in that Nozick does not provide an argument for the view that an individual is entitled to his natural assets; rather, Nozick argues against Rawls and assumes that entitlement is the correct alternative. Thus Nozick establishes a system on weak grounds that ultimately maintains a level of arbitrary rewarding of certain individuals and does not recognize the need for its reform. In contrast to Nozick, Scheffler’s conception does not ignore this point and at least acknowledges the plight of the naturally less advantaged individuals and attempts to remedy their situation. This argument, however, does not address the objection to Scheffler regarding treating individuals as means, rather it merely provides one point where Scheffler’s view can overcome a difficulty with which Nozick’s view can be confronted.

Thus far, I have presented an objection to Scheffler’s alternative theory of rights and provided points of clarification, two of which could work in Scheffler’s favor, the other ultimately reducing the objection to a fundamental question of competing priorities. These points, however, take for granted that the premises of the objections are correct; they assume that either taxation is the method for providing the needy with the means for living a meaningful life, or alternatively, that well-off individuals must be used as means for the ends of the aiding the less advantaged. Scheffler does not have to grant either of these claims. Moreover, a careful and strict reading of his view will demonstrate that his alternative specifies a stricter proviso than Nozick’s and limits to a greater extent the amount of goods to be obtained by individuals. In contrast to Nozick’s proviso, which merely stipulates that any appropriation cannot leave anyone worse-off, Scheffler’s alternative stipulates that no one can obtain any good that prevents someone else from living a meaningful life. Accordingly, any appropriations that may end-up preventing someone from living a meaningful life are prohibited. The state thus actively provides
positive rights to individuals by ensuring that the state of affairs in society will not result in a situation where certain individuals have more than they need at the expense of others not being able to fulfill their capacity to live a meaningful life. Nozick is not concerned with ensuring that individuals have the capacity to live a meaningful life, rather he is interested in maintaining a minimal state that promotes a system of free exchange.

While Scheffler may accept the re-reading of his claim as presented above, I am not certain that he would capitulate on providing such positive rights as welfare, education, and other social programs, and probably not object to using taxation as means for their implementation. The question of prioritizing the prevention of treating individuals as means over ensuring that individuals fulfill their capacity to lead a meaningful life still remains. In this regard, it might be tempting to utilize Rawls’s comment on the nature of philosophical problems:

. . . these remarks suggest that in philosophy questions at the most fundamental level are not usually settled by conclusive argument. What is obvious to some persons and accepted as a basic idea is unintelligible to others. The way to resolve the matter is to consider after due reflection which view, when fully worked out, offers the most coherent and convincing account. About this, of course, judgments may differ.\textsuperscript{15}

Were we to take Rawls’s quote to its fullest extent concerning this conjecture we would be left at a standstill. Clearly, the notion of a meaningful life plays an integral role in this problem, and indeed, the fundamental question above can be understood in this respect. Both Nozick and Scheffler agree that the capacity to live a meaningful life ultimately forms the foundation for their conception of rights: for Nozick, the capacity to live a meaningful life begins at birth and obtains in virtue of being human; for Scheffler, it seems that the capacity to live a meaningful life is effectively meaningless without its actualization. Based on what has been presented in \textit{Anarchy, State, and Utopia}, Nozick has yet to establish his case regarding the notion of meaningful life. Moreover, Scheffler has shown that a different conception of rights (at least in some form) can be substantiated even within Nozick’s interpretation of a meaningful life. While this notion of a meaningful life and its proper place in the debate regarding the form of rights still remains elusive, intuitively one should still be bothered by the idea that Nozick’s theory refuses to assume any responsibility for establishing a state of af-
fairs that condones the undeserved rotten lot of some when the means to improve it are known.

NOTES

♦ I am profoundly indebted to Carla Yumatle, whose insightful comments and continuous aid greatly influenced the content of this paper.
2 Here Nozick primarily considers classic Utilitarianism, where the total net utility is the goal. The theory of average utility would also be subject to these criticisms, as would all end-state theories.
3 Nozick derives this conception of side-constraints from Kant’s *Groundwork of the Metaphysics of Morals*.
4 Nozick re-formulates Locke’s Proviso “leave some and as good for others,” into a more lax proviso, specifying that only appropriation that worsen others’ situations are not permissible.
5 As a consequence of the entitlement conception of justice, Nozick considers taxation by the state unjust if it serves any purpose other than to ensure the protection of individual rights. Nozick regards taxation on labor tantamount to forced labor; see Nozick pg. 169.
6 Nozick regards involuntary actions as those that i) the actions of an individual are restricted, and ii) this restriction brings about the violation of the individual’s rights. Note that this definition is consistent with the second clause of the first principle of justice, i.e., no one’s situation worsens as a result of acquisition, “or leave some and as good for others.” In acquisition, there is a restriction of action, but this restriction is considered legitimate.
8 The base-line favored by Nozick, “common ownership” as it is referred to by Cohen, considers how someone would have fared had no private appropriation taken place and had as a result the object of appropriation been left available for the free use by everybody without anyone’s appropriation of it.
10 These two claims are presented by Alan H. Goldman in “The Entitlement Theory of Distributive Justice,” *The Journal of Philosophy*, volume 73, issue
ENTITLEMENT, OWNERSHIP, MEANINGFUL LIFE

11 See David Conway’s “Nozick’s Entitlement Theory of Justice: Three Critics Answered,” in Philosophical Notes no. 15, (1990: Libertarian Alliance; David Conway), and see also, Jonathan Wolff’s, Robert Nozick: Property, justice, and the Minimal State (Polity Press, 1991), pg 32.

12 The capacity to live a meaningful life is not to be confused with the actualization of one’s living a meaningful life. Nozick contends that the capacity for an individual to have a meaningful life may be one of the defining characteristics of a human being. Thus all human beings should not be treated as means in virtue of having this capacity as a human being to live meaningful life.

13 Arthur Flemming, “Using a Man as a Means,” in Ethics, volume 88, issue 4 (July, 1978), 283-298. In relating his argument to Nozick, Flemming is bothered by Nozick’s analogy of a tool used to explain side-constraints. Nozick explains that side-constraints are limitations on actions, such as would be applied to a tool not just to preserve it for future use (e.g. “don’t leave it out in the rain”), but also as an overridable constraint C on some tool’s use. In this case, “the object may not be used as a tool in those ways. In those respects, it is not a tool at all” (Nozick, 31). Flemming states that it is not clear what the tool analogy is designed to prove. Despite the overriding constraints, the object is still a tool and can be used as such, provided that the constraints are not violated. The same is the case for many objects that are used in life. Thus it seems that side constraints can be viewed as impositions on specific kinds and results (i.e. damage) of use and not as constraints denying all use. In its development, Flemming’s argument concludes that there are some ends for which individuals can be used as means as long as certain conditions on their use (e.g. avoiding injury, etc.) are met.


RESPONSE BY: ADAM SASSOON

In *Anarchy, State, and Utopia*, Robert Nozick argues for a minimalist state that serves only to (a) protect the negative rights of its citizens and (b) rectify violations of their rights\(^1\). Nozick’s emphasis on the individual citizen is central to the entitlement theory of justice and needs to be considered in any criticism of that theory. In this paper, I will briefly explain Nozick’s view on the natural rights of man, roughly outline G.A. Cohen’s criticism against Nozick’s entitlement theory, and point out that Cohen’s idea of joint ownership fails to present a sufficient alternative to Nozick’s baseline of common ownership. I will illustrate this point with a case of extreme disagreement in Cohen’s two person world and argue that joint ownership is insensitive to such cases because they may lead to infringements of negative rights. Furthermore, I will attempt to show that these cases can also result in situations where individuals are made worse off through joint ownership than they would be when compared to Cohen’s “actual situation.” I will then go on to present some possible replies that Cohen could offer to these claims.

**Nozick’s Conception of Rights:**

Nozick’s justification of a minimal state stems from two ideas: (1) man is entitled to certain natural rights and, (2) these rights should take the form of side constraints (Nozick, 31). According to Nozick the natural rights of man are negative rights, or rights of non-interference. Examples of these negative rights include a man having the right to *not be* beaten, coerced, etc. The fact that these rights take the form of side constraints, means that they may not be violated, even in order to prevent future violations of rights. Nozick’s view of negative rights stems from the Kantian ideal that man should be seen only as an end and never as a means.

**Cohen’s Plan of Action:**

In order to develop a strong argument against Nozick, Cohen works within Nozick’s theory of negative rights to show that, even granting Nozick these fundamental claims, Nozick’s choice of common ownership as a baseline for his entitlement theory is arbitrary. Cohen grounds his criticism against Nozick’s entitlement theory in a weak version of self ownership. In this version, Cohen accepts the fact that an individual owns his body, but he is not necessarily entitled to the products of his talents. Cohen attempts to promote a thesis for this form of self-ownership through his example of an “eye lottery.”\(^2\) In this example, Cohen imagines a world where eye
transplant surgeries are commonplace, and individuals with two eyes are conscripted by the state to surrender an eye to individuals who are blind in both of their eyes. This situation appeals to an intuition that, even though people do not deserve two functional eyes, and that they do not need both of them more than a blind person would need one, their claim to their eyes is not weaker than some unfortunate blind individual. The “eye lottery” situation not only highlights Cohen’s affirmation of self-ownership, but also his affirmation of negative rights, in that people have the right to not have their body interfered with.

Cohen states that in spite of his affirmation of self-ownership and negative rights, the drastically unequal distribution of external goods resulting from the entitlement theory, which Nozick develops from self-ownership, is unwarranted because the entitlement theory results from an arbitrary basis of comparison (Cohen, 72). Cohen uses his two person world to show that other baseline scenarios, or pre-appropriation situations, can be used which put inhabitants in a better position when compared to common ownership. By showing this, Cohen claims that post-appropriation situations that would not violate Nozick’s proviso using a common ownership baseline, will break the proviso, by making one of the inhabitants worse off, if an alternative pre-appropriation scenario operates as a baseline. Cohen presents joint ownership as an example of one of these possible baseline situations.

**Forms of Ownership:**

Before launching a criticism against Cohen’s choice of an alternative baseline scenario, it is crucial to highlight the key distinction between joint ownership and common ownership. Common ownership affords each individual the right to use the land without the consent of other individuals if this use does not interfere with another individual’s use of the land. Joint ownership requires the consent of every individual before a specific individual can use the land (Cohen, 84).

**Problems with Joint Ownership:**

While Cohen attempts to work within Nozick’s theory of negative rights, or rights of non-interference, his offer of joint ownership as an alternative baseline situation to common ownership can be made inconsistent with negative rights in certain extreme situations. Joint ownership places every individual in a situation where he is able to exercise power over the rights of every other individual. Consider a variation of Cohen’s two person world in which A bitterly despises B to the point where he is willing to endure starving himself to death so that B will die as well. Joint ownership affords A this
possibility by allowing him to veto any of B’s proposals for the use of the jointly owned land. Even if, in other variations of the two person world, joint ownership would allow for a situation where A and B are better off when compared to the actual position, it also allows for one’s negative rights to be violated. This is unacceptable for Nozick who regards negative rights as side constraints. Cohen could reply to this claim by stating that the weight then falls on Nozick to prove that negative rights should take the form of side constraints and uses this argument against Nozick’s Wilt Chamberlain example. Cohen states “side constraints remain unjustified, and socialists need not apologize for being willing to restrict freedom in order to expand it.” However this line of reasoning does not apply to Cohen’s two person world as smoothly as it does in the case of Wilt “the Stilt” Chamberlain.

In Cohen’s two person world, freedom is not being restricted to expand it, rather it is used to harm another individual. This matter of whether or not side constraints are justified is one that will not be sifted through in this response; however, Cohen’s inability to make his positive claims about joint ownership consistent with Nozick’s idea of side-constraints, weakens his overall argument, that common ownership is arbitrary. It seems that Cohen cannot arrive at joint ownership, as a viable situation, if he grants Nozick all of his fundamental ideals.

Taking a slightly milder variation of Cohen’s two person world in which A hates B to a lesser degree, some other problems with joint ownership can be brought to the surface. In this example A’s hatred for B has waned so that he is unwilling to take his own life to see B die. However, A is willing to frustrate B to the point of near starvation. In this scenario A vetoes all of B’s proposals to the point where they are both about to die; only then does A consent to B’s proposal to harvest or pick roots etc. Once A escapes death, the cycle begins again so that both A and B are living in a state of continual suffering and malnutrition. If you consider this state of affairs over a given period of time, a month for example, it seems unlikely that A and B are going to be better off under joint ownership than they would be in Cohen’s actual situation or even a state of common ownership. If B did not have to gain A’s consent before using land then he would be more productive, using the land on a daily basis rather than once every five days in a state of exhaustion. Furthermore, A would also be more productive. Knowing that he could not prevent B’s use of commonly or privately owned land, he would give up his fancies of self-denial and be more productive as well. Cohen could reply that this situation does not recognize B’s ability to veto A’s decision which allows the land to go unused, however, A could go on to veto B’s veto so that the situation degenerates into a ping-pong like rally
of vetoes. It seems that this would result in the land remaining unused due to the inability of A and B to reach an agreement. Thus functionally, B’s veto of A’s decision to not use the land is impotent and B’s consent is really not necessary for A to carry out the action of inaction. If you expand this scenario to a multiperson world then it seems that one person could cause a whole group of people to live in constant suffering, however, I do not believe that Cohen would want to allow for this. If the land is used with the consent of everybody minus the consent of one individual, through a vote or some other form of decision making, then we are no longer dealing with joint ownership strictly defined. Cohen recognizes this problem:

…the land is owned, by all together, and what each may do with it is subject to collective decision. The appropriate procedure for reaching this decision may be hard to define, but it will certainly not be open to any one of the joint owners to privatize all or part of the asset unilaterally, no matter what compensation he offers to the rest.\footnote{\textsuperscript{4}}

However, his recognition that this is a problem does provide a solution, for it seems plausible that situations can arise in which the actual situation satisfies Nozick’s proviso with joint ownership acting as the baseline of comparison. This shows that joint ownership is subject to the same type of criticism Cohen uses against Nozick’s baseline of common ownership, in that identical appropriations can be rendered just or unjust on the basis of outside circumstances.

While these variations of Cohen’s two person world are examples of extreme disagreement, they bring out the facts that joint ownership could allow for the violation of negative rights and that joint ownership does not always guarantee that individuals are put in a better situation when compared to common or private ownership. These points, however, do not take away from Cohen’s claim that it is still up to Nozick to prove that negative rights should act as side constraints. Furthermore, the flaws in joint ownership which were highlighted by my examples, do not show that the rest of Cohen’s argument against Nozick’s theory is ineffectual. Rather, the cases discussed in Noam Jacob’s paper, in which Cohen compares the actual situation to situations in which A and B are equally talented or in which B is more gifted than A, still pose a problem to the strength of Nozick’s proviso.
NOTES

♦ I would like to express my warmest appreciation and extreme gratitude to Carla Yumatle for her valuable comments, guiding words, and constant encouragement. This response would not have been possible without them.

The purpose of this response is to provide a rhetorical understanding of the issues surrounding political justice as argued by Samuel Scheffler, David Conway, and Robert Nozick. By rhetorical approach, I do not mean trying to understand the persuasions of the arguments provided, rather, trying to understand what assumptions the previous discourse must take for granted before beginning the debate. In *Anarchy, State, and Utopia*, Robert Nozick presents his entitlement conception of justice beginning the discussion at hand. In “Natural Rights, Equality, and the Minimal State” Samuel Scheffler interprets a section of Nozick’s work to support an “alternative conception of rights,” better than Nozick’s’ own conception of rights. After explicating this move I will present a criticism of Scheffler provided by David Conway, and a possible response Scheffler could provide to Conway. This is precisely what is taken up in part by Noam Jacob in his paper. His essay takes a critical stance of Nozick via two criticisms provided by G. A. Cohen and Scheffler. What I propose, however, is neither a criticism nor an endorsement of either philosophical view. Instead, the purpose of this response is to examine the aspects of political philosophy on which they base their dialectic. I hope to show that because the original argument provided by Nozick is based on Kantian moral claims the discussion itself cannot lead to any final conclusions as to what form justice should take. These Kantian claims are themselves very complex, but can be understood as a person’s right to live a meaningful life in which he is only used as an ends and never a means. Finally, I hope to summarize an alternative theory of justice outlined by John Rawls in *Political Liberalism* which is in turn not rooted in comprehensive moral claims. This is not to purport that the previous philosophers were wrong in basing their conception of political justice on these grounds, only that moral grounds for political justice can result in equally just, but conflicting claims about which form justice should take, and to point out that other grounds for political justice are possible. Noam began to flush this issue out at the end of his essay as “the nature of philosophical problems,” and in part, this response is an attempt to understand what that problem is in relation to arguments of political philosophy.

*How Nozick Arrives at Kantian Morals*

Nozick claims that the only rights people have are negative rights that
take the form of side-constraints. Thus Nozick describes a just state as one that uses its power only to protect the negative rights of its citizens, and punish those who threaten them. One important aspect of this is the right to accumulate and maintain property, provided that it was acquired by a just means. What separates Nozick’s view from others is the belief that the protection of these rights constitute the full extent of a government’s power. Nozick does not believe that people have any positive rights. Under Nozick’s interpretation of rights, it would be unfair for a government to tax its citizens in order to provide public buildings or hospitals, or any other public service besides the police and the military (who protect the negative rights of citizens). To tax citizens in order to provide positive rights would be similar to the state stealing its citizen’s properties. This is because the negative rights of a person protect them from having any property taken away from them without just cause.

One criticism of Nozick’s entitlement theory is that it seems to require a state of unlimited capitalism. Any society that prohibits the exchange of goods between consenting adults would be considered tyrannical. An unfortunate aspect of this historical conception of distribution is that it is very possible to imagine situations in which there is a gap between those who have acquired large amounts of property and those who lack enough property to live in comfort. Despite the initial shock of such a theory of distribution, Nozick asserts that such differences in outcome still adhere to justice. The justification of this theory of distribution stems from Nozick’s conception of rights as side constraints, and so this raises the question of what side constraints are based on.

Nozick does not provide a clear account of why rights exist at all; instead he seems to presuppose them. The first sentence in his preface is “Individuals have rights, and there are things no person or group may do to them without violating their rights” (Nozick, ix). Nozick does, however, elaborate on Kant’s moral claim that people should be treated as ends and not means. From this, Nozick believes that any being which can construct a certain life plan for itself (as an end) has a certain claim to natural rights (not to be treated as a means). According to Nozick, to construct a certain plan of life is what gives that life meaning. While this is the closest thing to an explanation of why people have rights, it seems then that rights rest on the ability of a person to construct a life-plan. If this is the case, then Nozick’s negative rights do not necessarily protect this ability. There are many situations in which a person’s ability to construct a life plan is threatened because society did not provide any positive rights. For example, most people construct a life-plan assuming they will stay healthy. For such a person, without
their health, they could not live a meaningful life. This would perhaps result in the moral claim, and thus the requirement to provide healthcare to a sick person. While taxation for hospitals would still be redistributing a healthy person’s money, and therefore violating his right to property, if rights conflict, I do not see why the property rights of one person would override the health rights of another.

Scheffler’s Account of Rights

The account of positive rights described above is similar to Scheffler’s position in “Natural Rights, Equality, and the Minimal State.” As Noam explains in section three of his essay, Scheffler interprets Nozick’s moral basis of negative rights (living a meaningful life), as supporting a more inclusive and “alternative view” of rights better than Nozick’s own entitlement conception of justice. Any person who has the capacity to live a meaningful life has the right to have that capacity protected. This account of positive rights is not only more egalitarian and in my opinion more appealing, it also seems to be rooted on the moral claim that people should be treated as ends and not means. In this case, a person who is unable to live a meaningful life, is unable to accomplish any ends. Thus, they must be supported by the state so that they will at least have the capacity to live a meaningful life and accomplish some kind of ends. However, it is also easy to see where this kind of argument is subject to criticism from a Nozickian response. This response as I imagine it would appeal to the claim that in order to redistribute goods to those who need them so that they have a meaningful life, others must be treated as a means through the taxation of their property for the benefit of others. In respects to the second party, such taxation according to Nozick does not take seriously the fact that “his is the only life he has.” (Nozick, 33) Such an argument is made by David Conway in “Nozick’s Entitlement Theory of Justice: Three Critics Answered.” In this response to Scheffler, Conway addresses what I have earlier eluded. “…if rights conflict, I do not see why the property rights of one person would override the health rights of another.”

Conway characterizes Scheffler’s account of rights as the justification of having one’s needs met “purely by virtue of their need[s].” (Conway, 5) While he admits this is an appealing view, Conway makes the claim that the right of one person to their property is not morally equivalent with the right of another person to have their basic needs met. While Scheffler finds the justification for his “alternative theory of rights” from a section of Nozick’s
work called “What are side constraints based on?” (Nozick, 48-51) Conway appeals to the section titled “Why side constraints?” (Nozick, 30-33). While the former section explains that a persons right to lead a meaningful life must be protected (resulting in Scheffler’s “alternative rights”), the latter section supports the claim that a person cannot be taxed in such a manner described by Scheffler, as this violates his right to lead an individual life. Simply stated, this is the claim that a person may not be used as a means towards the ends of another without their consent. Since for Nozick positive rights are unjust, the kind of taxation eluded to by Scheffler would violate the previous claim and would thus be deemed unjust. Conway grants Scheffler the supposition that positive rights do exist, and still argues that Scheffler’s claim for the protection of the capacity to live a meaningful life does not justify the taxation of individuals’ property, as it would still violate the rights of those people to their property. While the problem of evaluating these contrary morally claims seems difficult, Conway believes that Scheffler’s “alternative rights” are not moral equivalent to Nozick’s side-constraints. He does this through a thought experiment including two people A and B, one of whom is permanently disabled.

Person A’s being compelled to refrain from harming person B does not exact from A any greater sacrifice than is exacted from B for A’s benefit when B is likewise compelled to refrain from harming A (Conway, 7).

Contrary to this, Conway states that person A is required to make a larger sacrifice for B’s benefit than is exacted from B for A’s benefit where B is permanently disabled. From this, Conway concludes that Nozick’s entitlement conception stays more true to the claim that people should not be used as a means.

Scheffler’s reply to Conway would probably appeal to what constitutes rights in the first place and what it is for an individual to have rights. Just by the mere fact of having the capacity to live, a person has certain rights that cannot be taken away from them. These rights are the same for all people in virtue of all people being able to create a certain life plan. A life plan in turn becomes a meaningful life. According to Nozick a meaningful life should be protected, resulting in negative rights which act as side-constraints, but I think Scheffler would stress that it is the capacity to create a life plan that should be protected, and not the details of the plan itself. If this is the case, then a person should be equipped with whatever goods are necessary to have the capacity to live a meaningful life. If natural rights are based on
the ability (and not the actuality) of living a meaningful life, then B’s right to being compensated for his permanent disability is not overshadowed just because it redistributes a certain amount of wealth from A. A’s ability to imagine and therefore live a meaningful life is not at all threatened by such a redistribution, and so his overall rights are not violated. This point brings up a very important question. What if this redistribution of wealth does make it impossible for A to live a meaningful life?

Why Scheffler’s Account of Rights is Based on Moral Grounds

Scheffler accounts for such a situation in the proviso of his “alternative conception of rights.” Scheffler states that

Every person has a natural right to a sufficient share of every distributable good [which]...is a necessary condition of the person’s having...a decent and fulfilling life, subject only to the following qualification. No person has a natural right to any good which can only be obtained by preventing someone else from having a decent and fulfilling life (Scheffler, 153).

Scheffler’s original justification for a redistribution of wealth was to protect a person’s right to have a meaningful life. The right to have a meaningful life is in turn justified by the concept that people are ends and not means. In this sense, Scheffler does not want to throw away the concept of using someone as a means, but only re-defines what it is to use someone as a means in a broader manner. To deny A his basic needs in order to provide B with his basic needs would be using person A as a means, and this is why this sort of redistribution is impermissible. As long as A’s basic needs are not threatened by a redistribution, then his capacity to have a meaningful life, and in turn his ability to be an end is not violated.

A Re-evaluation of the Discussion

While Conway and Scheffler interpret Nozick in different ways, each in responding to Nozick base their theory of justice on Kant’s moral claims. Nozick and Conway have a more strict understanding in which any forced redistribution of wealth uses a person as a means, and therefore any state
enforced redistribution of wealth is unjust. Under this model of justice, a person can live a meaningful life knowing that his rights to property will be protected. Scheffler has a broader interpretation of this claim that people should not be used as a means. This interpretation finds justification by appealing to the claim that people cannot be treated as means because of their ability to form a rational plan of life. This ability then becomes the final criterion of whether a person is being used as a means. If this ability is taken away, then a person is being used as a means, but if this ability remains, then the person in question is able to live a meaningful life knowing that his right to this capacity will be protected. Either way, the end goal of these interpretations is to ensure that people are not used as means. The difference then between Nozick, Conway, and Scheffler is not one of kind, but the extent to which the same principles can be invoked. While it is possible to find more textual evidence in support of any of these philosophers, it is difficult to imagine a way to completely refute Nozick’s claim that individuals should not be used without their consent, or Scheffler’s claim, that the ability to have a meaningful life should be protected. I believe this is a symptom of basing their discussion on a moral claim.

Rawls’s Conception of Liberal Theory

The question of whether or not a moral claim is the best ground to begin a discussion of political justice still remains. While it might seem very intuitive to begin a discussion of political justice by appealing to normative claims, it is easy to see how such a grounding can result in continued debate and confusion. The previous analysis does not even touch on competing moral claims outside of philosophy in areas such as religion, which would only further complicate matters. In Political Liberalism John Rawls takes up this very issue.

In his earlier work, A Theory of Justice, Rawls sets up a thought experiment in which parties are placed in the “original position,” in order to commit to a theory of justice. In this position, the parties are placed under a “veil of ignorance” so that no individual has any knowledge about who they are or what they believe. This constraint not only makes it impossible for anyone to set up a system of distribution which would be in their favor, but also forces a certain kind of consensus. In this consensus the following liberal ideals are agreed to because, as seen from the parties in the original position, they would be to everyone’s benefit.
First: each person is to have an equal right to the most extensive scheme of equal basic liberties compatible with a similar scheme of liberties for others. Second: social and economic inequalities are to be arranged so that they are both (a) reasonably expected to be to everyone’s advantage, and (b) attached to positions and offices open to all (Rawls (a), 53).

It is easy to imagine why these principles, which make up “justice as fairness,” would be agreed to by the parties in the original position, who do not know anything about themselves. In *A Theory of Justice*, however, Rawls does not seem to answer the question of what would occur once this “veil of ignorance” is removed.

The central question in *Political Liberalism* is why people who share different comprehensive moral doctrines could agree on any one theory of justice. It is obvious from the discussion above and Noam’s essay that Scheffler and Nozick do not agree on a theory of justice, and they are working with the same comprehensive moral doctrine. It appears then that it is highly unlikely that there will be consensus on any one moral doctrine in a pluralistic society. This idea leads to Rawls’s belief that his theory of justice, which is not based on any moral claim, is the only liberal theory that can create a stable and free society. Nozick and Scheffler each provide a liberal theory of justice insofar as they deem the fundamental role of society to be the protection of the individual. This protection is characterized differently by each philosopher due to differences in their understanding of the same comprehensive moral doctrine. In *Political Liberalism* Rawls shows that since a theory of liberalism must take into account the various moral beliefs that individuals have, a political conception of liberalism cannot itself be based on any one moral claim.

**Rawls Escapes Moral Arguments**

Rawls’s conception of justice is considered to be de-teleological, because an understanding of the good does not define justice. An example of a teleological theory would deem something as morally good, such as “not using a person as a means” and then define what justice is in relation to that good. Instead, Rawls believes that a theory of justice must define the limits of justice, and allow for any person to achieve any good within those limits. Any person can live their life in any way as long as it adheres to the sense
of justice agreed to by society. Rawls creates such a theory of justice in part because he believes that choosing any ‘comprehensive moral doctrine’ as the grounds for a theory of justice does not take seriously the claim that individuals can live their lives according to their own beliefs.

*Political Liberalism* is based on the political conception of society as a system of cooperation among free and equal persons for mutual benefit. Because of this foundation, Rawls is able to justify any society that remains stable and continues to uphold the liberty and equality of its citizens. Rawls believes, however, that a stable society that upholds liberty and equality must allow for individuals to have their own private comprehensive moral doctrines. While it is expected that these private moral doctrines will conflict, Rawls argues that a public conception of justice will be based on the shared aspects of these private moral beliefs. People are allowed to disagree with each other about moral claims because Rawls states that such disagreements are “reasonable” and even expected due to the “burdens of judgment” which are placed on all citizens in choosing a moral doctrine. Society’s sense of justice is also achieved through a sense of reasonableness, by taking into account where the majority of different moral doctrines overlap and hence where reasonable people can agree to form a political conception of justice. This point of overlapping moral doctrines becomes the basis for a public conception of justice. As it stands, a public conception of justice is not itself a comprehensive moral doctrine, but rather the point at which many different moral doctrines overlap. While citizens must conduct themselves in a manner that adheres to the limits of the public conception of justice, they can privately conduct themselves in any manner they see fit. I believe examples of this can be seen in our own society concerning nudity and drunkenness. These actions are impermissible in a public forum because many reasonable comprehensive moral doctrines do not provide room for such acts in their belief structure. As a result, the public conception of justice does not permit these kinds of actions in a public setting. This public conception of justice, however, is not a moral doctrine itself and so these actions are not ruled as “unjust,” only restricted in a public stage. Thus, citizens are allowed to pursue these ends in private. Another more pertinent example of Rawls’s point developed in *Political Liberalism* can be seen when taking into consideration our separation of Church and State. The state cannot be grounded in any one religious moral doctrine, but does not take away the ability of individuals to pursue such moral doctrines in a private manner.

It is easy to see how *Political Liberalism* and *A Theory of Justice* work together to create Rawls’s overall liberal theory of justice. The individual is protected from the authoritarian rule of society not because of a moral
claim, but because in the “original position,” where the institutions of society are decided, the parties would have accommodated for all possible comprehensive moral doctrines to which they could reasonably agree. This society is stabilized by creating a public conception of justice that, although is not itself a comprehensive moral doctrine, is constructed from the points at which different comprehensive moral doctrines overlap. It is possible to imagine that without this political conception of justice, individuals who hold different comprehensive moral doctrines would be unable to live in the same society, as each would constantly be accusing the other of committing injustices according to their moral doctrine. This possible debate would appear to have no definite final solution, as Noam put it, the nature of these “philosophical problems” are never easily sequestered.

Back to the Argument with a New Perspective

The discourse of political justice between Nozick, Scheffler, and Conway is a very intricate and important aspect of political philosophy. As a result of Nozick beginning this discourse, Scheffler and Conway are limited in their ability to respond without appealing to moral claims. The purpose of this response is to show that by grounding their arguments in moral claims, no final conclusion can be made as to what form rights should take. Both Scheffler’s and Nozick’s arguments find moral support in the same comprehensive moral doctrine. Despite Conway’s objections to Scheffler and his possible response, neither provides an irrefutable argument as to which form the Kantian moral claim should take. “...in philosophy questions at the most fundamental level are not usually settled by conclusive argument. What is obvious to some persons and accepted as a basic idea is unintelligible to others” (Rawls (b), 53). Rawls on the other hand avoids this difficulty by bypassing moral claims all together. His theory of justice depends on peoples’ ability to reasonably agree, thus reaching a stable and equal society. Agreement of this kind is not grounded in any one moral doctrine (not even the doctrine that agreement is good), but relies on a plurality of moral doctrines to begin forming a political conception of justice. I think there is a strong sense in which this theory is intuitively appealing. If we are to take seriously the liberal claim that individuals have the right to live their life as they see fit, then we cannot define justice in relation to a moral doctrine that is particular to only a few of those individuals.

Thus, a main aim of ‘Political Liberalism’ is to show that
the idea of the well ordered society in ‘A Theory of Justice’ may be reformulated so as to take account [for] the fact of reasonable pluralism. To do this it transforms the doctrine of ‘justice as fairness’...into a political conception of justice that applies to the basic structure of society. [In footnote: By the ‘basic structure’ is meant society’s main political, constitutional, social and economic institutions and how they fit together to form a unified scheme of social cooperation....This structure lies entirely within the domain of the political.] (Rawls (b), xiii)

Political conceptions of justice in which the application of public reason is limited in scope and context results in a society that is able to remain stable, and allows for the citizens of that society to remain free and equal. Political Liberalism is able to achieve this goal by understanding the intricacies of individuals and pluralism. This is its continuous theme, and it is an important concept that I believe should be invoked when discussing any theory of political philosophy.

NOTES

1I am grateful to Adam Sassoon for invaluable discussion of these points.

WORKS CITED

In what follows, I will i.) press a line of argument that has been but grazed by the discussion of a person’s “capacity to lead a meaningful life” thus far: namely the role self-ownership plays in discussions of liberty and equality ii.) explain how the concept of self-ownership is the core notion of libertarian thought, and how it drives the rejection of patterned principles of distributive justice, iii.) briefly assess the discussion thus far, and iv.) argue that Cohen’s arguments against the strong thesis of self-ownership need be taken seriously by egalitarian thought generally, and not just in the context of answering the libertarian challenge.

1. Liberty and Equality?

What is involved in the claim: “people have the capacity to lead meaningful lives”? Well first, note that the claim makes some assertion about the sort of beings we are, that we do in fact have certain a capacity. Second, note that separable questions of need and entitlement will arise when determining what that capacity amounts to: are certain items such as food and clothing, or shelter and work, to be regarded as basic human needs? Also, given what criteria should we say that certain persons are entitled to certain service and product? (I shall of course have much more to say about what may be involved in the latter point shortly.) Third, and most importantly, underlying the claim is the idea that if people do in fact have that capacity, a given mode of social organization should be responsive to it, however we characterize it. The question I would like to pursue here, is then this: Given a certain conception of what people are, does it follow that the state, or any of its institutions, are required to provide those resources we might regard as necessary for the exercise of the capacities they have?

Liberal egalitarian theories before and after Nozick have answered yes: it is incumbent upon the state to provide some share of those goods deemed necessary to make it at least possible for individuals to exercise that capacity. One famous proposal by Prof. Williams, is that medical care, or the vital goods that remedy ill-health or emergency, should be based on an individual’s need, and not whether she can afford to pay for those goods or services. Another is that of Ronald Dworkin, who has argued that in order for a state to legitimately claim the right to govern, it must distribute goods equally among its citizens, even if that involves abrogating certain rights or
claims to liberty. There is a trade-off between a person’s entitlements, which negative rights are designed to protect, and other person’s claims of need, which involve the potentially problematic notion of positive rights.

Nozick would presumably have to respond no: it simply is not justifiable for the state to distribute the goods of some for the general welfare of others. Why not? Given Noam’s explication of the entitlement theory, the answer should be clear: Individual rights have a certain inviolable status. These entirely negative rights are central to the entitlement conception of justice, these rights are pinpoint the area around which a theory of justice should be constructed: the protected sphere of individual liberty. So, even if redistributing wealth or other goods would enable all persons, or perhaps just the vast majority of persons, the resources necessary for the possibility of exercising that capacity, such a procedure could not be undertaken if it would violate rights to non-interference. As Nozick is famous for saying, “Taxation is on a par with forced labor” and “taking the earning of n hours labour is like taking n hours from the life of a person.” The individual, and those rights taken to be inalienably hers, such as the right not to have her earnings taken from her, simply cannot be violated.

A common objection to the theory is to ask why entitlement and individual liberty should be of sole importance to a theory of justice. The ability to live one’s own life free from unwarranted interference, as well as having one’s options open to pursue various projects, is something few would reject. But surely there are other principles and values: need, desert, perhaps efficiency, well-being, equality, etc., that should figure into the construction of a theory of justice. I shall not press those arguments, but shall leave open the question of whether or not individuals do in fact have certain inviolable, or “original” rights. If they do, then I assume that regardless of their form, those rights are not only necessary features for a theory of justice, but perhaps they should be taken as central to the construction of that theory. What interests me more is the claim that persons in a society have no non-contractual obligations to one another. On the face of it, the idea is that in a given society where rights are merely formal restrictions on the actions of others, there is nothing to compel me to act in such a way that might benefit others. If that is so, then the state cannot legislate, and thereby enforce by rule of law, distributive principles which would violate those fundamental rights. And if that is the case, then individual liberties will always trump efforts to equalize accidents of birth, respond to basic human needs, distribute resources, etc. Carried out to its furthest logical conclusion, individual liberty not only allows, and in a sense justifies, massive inequality, it rules out the implementing of any kind of egalitarian program. Freedom
and equality then do conflict.

2. Self-ownership: Understanding the libertarian position.

As Nozick says about distributive principles, for instance that the state should provide adequate medical care to those who need it: “[they] institute (partial) ownership by others of people and their actions and labor. These principles involve a shift from the classical liberals’ notion of self-ownership to a notion of (partial) property rights in other people.” Also, “…patterned principles of distributive justice involve appropriating the actions of other persons. Seizing the results of someone’s labor is equivalent to seizing hours from him and directing him to carry on various activities. If people force you to do certain work, or unrewarded work, for a certain period of time, they decide what you are to do and what purposes your work is to serve apart from your decisions. This process whereby they take this decision from you makes them a part-owner of you; it gives them a property right in you.” If Nozick is correct in these claims, we have the following consequences:

(1) Persons are in some relevant sense “self-owners,”
(2) If person X is under some non-contractual, (non-consensual?), obligation O for person Y, then the benefits that come from completing or fulfilling O, rightfully belong to Y. As such, Y is the (partial) owner, via O, of X,
(3) Persons can “own,” i.e have (partial) property rights in, other persons,
(4) For all quantities P of work done by X that is coerced or ill-gotten (unrewarded) by Y, P belongs to Y, and Y is the (partial) owner X in virtue of owning P,
(5) All principles enforced by the state that impose non-contractual obligations on persons, entail (1)—(4).

Regarding (1), what counts as the relative sense, that is, what would make (1) true? If the question is considered completely outside of natural rights theory, it would presumably be some sort of assignment of status or right to persons generally, namely that X is in command of herself, and has full right (barring of course cases where others’ rights are violated by X’s exercise of, or failure to forbear, a certain right.) to her own person. Note that this right does differ substantially from a right to non-interference. Non-interference rights simply state that persons are not to be harmed or
obstructed from doing whatever it is they choose to do. The right to self-ownership involves the entitlement of O for X, and justification for X to arbitrate with Y what X should be justly compensated for O.

Regarding (2) and (3), again considered completely outside natural rights theory, there are the following questions. The first is to ask whether X, when O is completed or fulfilled, has actually given up anything of herself to Y. We could hold that she has undertaken the labor, or provided the material goods, for the realization of O.

We might stipulate principles of justice, as does Nozick, that state only contractual arrangements incur actual obligations to perform or fulfill O, buy that might involve denying that (2) and (3) are ever possible. For the correlative right for X to arbitrate with Y over the O that is rightfully hers, the right of ownership to the benefits of O could be stipulated to be the right of the producer solely. There is of course the question whether a non-contractual O need be understood as non-consensual. For it part of the principles of justice that O is stipulated to be justifiable for X to perform or fulfill if there are intervening considerations about Y—need, welfare, etc.—that are taken to be Y’s natural rights, in which case it might be stipulated which rights have priority in that society. Such stipulations would bar, or possibly make acceptable for X, (3). Without going that route, one might also obviate (3) via incentives for X to do O for Y that are non-contractual (it seems hard to imagine how these would not just be the inner satisfaction, or moral delight, exuberance over serving one’s state, etc., that would be internal to the motivations for X to do O without claiming (3) above). Of course, one could avoid the whole issue in (2) by stipulating that all such O’s that might be necessary are part of the “contract with the state,” and as such, Y is not some less fortunate person benefiting from X’s doing O, but Y is the state. That suggestion, however, paints a societal picture few would readily endorse.

Regarding (4), it is important to note that it is a variant on (2) that substitutes “unrewarded” (Nozick presumably has in mind some sort of coerced or non-contractual) work P for O. In any case, it seems to suppose ways in which the principles of justice may be constructed so that one can grant (1), but forestall the move from (4)—(3), or onto the conclusion stated in (5). There is question in (4) whether P really is equivalent to O, or how Y really could be said to possess X’s P, but if one equates not the work itself, but the material effects/benefits of P, the answer seems fairly straightforward. Then, the same sorts of questions that arise with (2) apply here. Concluding (5) from either the abhorrence of (3) (arrived at by moving from (1) to (2) or from (1) to (4),) is the real issue. This is the crux of the conflict between
liberty and equality introduced in section 1 above.

Now if one considers the passages on Nozick’s own ground, within natural rights theory, then it is not as easy to forestall (5). The notion of self-ownership is itself an intuitive one with great appeal: a person X has full rights over her person in the same way that she owns inanimate objects. I take it that no one would argue against having the sovereign right over their own persons in whichever way they see fit. For Nozick, the right to self-ownership is taken to be the guarantor of liberty: the individual’s “protected sphere” accrues from that individual’s ownership, or sovereign right over, their own persons. The self-ownership principle is on a par with liberty, if it simply isn’t liberty itself, which is the principle value for the entitlement theory. We do of course have a general sense that individual liberty, dissociated from further questions of responsibility and rational/volitional control, means that we have the choice in some ultimate sense over our own persons. However, it is crucial to recognize that the libertarian thesis of self-ownership goes much farther: First, this command over one’s own person takes the form of a property right in one’s own person. The thought is not just that one is taken to have the right to decide how am I in control over my own person, but that my own person belongs to me. But more importantly, not only do I, and only I, have complete and sovereign rights over my person, but because I do, I owe no service or product to anyone else. That is to say that (3) above is a violation of my negative rights. This is the strong principle of self-ownership, or the libertarian thesis of self-ownership.

At this point it may be useful to explicate how this aspect of self-ownership the entitlement theory is firmly based within Lockean natural rights tradition. That tradition of course affords individuals with very strong rights to life, liberty, and property, and all actions by the state are limited by the rights individuals have. I shall not answer the question of whether Nozick’s views can be evenly and cleanly cleaved off from Locke’s, I merely wish to show, by means of historical precedent, how one might move from the univocal rights over her own person to those of the stronger, more Nozickian flavor, namely that I those rights dissolve any obligations I may have to the state or others, and that the benefits of the work of my own hands should clearly belong to me, for I produced them. I hope that a quick presentation of Nozick’s indebtedness to the Locke on just-acquistion, appropriation, and self-ownership will suffice to make the point. The fundamental right around which the entitlement theory is based, presumably liberty, is rather liberty understood in an attenuated sense: that liberty to their possessions, which include their own persons, which is guaranteed to individuals by the
non-interference of the state. Property rights of the external world not only depend upon the property right one has over one’s own person, but that is how they are derived.

But what is the connection between having a right over one’s own person, however that might be construed, and rights over the material world? Do the sort of rights that one has over one’s person extend to those parts of the world—such as land, agriculture, buildings, and machinery—captured by notion ‘private property’? If so, how? As Locke famously claimed:

Every man has property in his own person: this nobody has any right to but himself. The labour of his body and the work of his hands, we may say, are properly his...For this labor being the unquestionable property of the laborer, no man but he can have a right to what that is once joined to, at least where there is enough and as good left in common for others.

...he who appropriates land to himself by his labour does not lessen but increase the common stock of mankind...he that encloses land, and has a greater plenty of the conveniences of life from ten acres than he could have had from a hundred left to nature, may truly be said to give ninety acres to mankind: for his labour now supplies him with provisions out of ten acres, which were but the product of a hundred lying in common.

Tis labour, then, which puts the greatest part of value upon land, without which it would scarcely be worth anything...Nature and the earth furnished only the most worthless materials, as in themselves. 12

As Nozick interprets the claim of right: “the central core of the notion of a property right in X is the right to determine what shall be done with X,” which is “the right to choose which of the constrained set of options concerning X shall be realized or attempted. The constraints are set by other principles or laws operating in the society; in our theory, by the Lockean rights people possess.”13 To take up the familiar claim that Noam has already mentioned: the principle that appropriation, whether it be in acquisition or in transfer, is a just situation if it arrived at by just steps14.” In the Lockean
conception, that is primarily through the effects of labor. The piece of land is yours because you found it first, you put the work in to cultivate it, and you have left ‘enough and as good for others.’ One moves from the thesis of self-ownership, to the just ownership of material goods in such scenarios, through just steps. One possible such step is the creation of value through labor, as in the second quote, leaves more and better for everyone else. For there is now cultivated land where before, (for Locke), there was unproductive and unprofitable wilderness. This is not to say that just-acquisition at the level of baselines need take this form, but that the ‘just situation by just steps’ proviso includes some direct connection between self-ownership and ownership of the external world. Note that this is not the only just step, for one could simply find an object, or make an artifact, etc., but as regards property such as land and agriculture, appropriation would require some extension of one’s own “protected sphere of rights.” That extension for Locke, and what Nozick is keen to preserve with the “taxation on a par with forced labor” “working for another is giving them a (partial) property right in you” rhetoric, is made through one’s labor. And this is why Nozick thinks that individual liberty precludes distributive principles of justice, for those principles are based on a sense of certain non-contractual obligations between individuals and the state. The individual, because she is a full self-owner, has no duties to the state or her fellow citizens beyond what is contractual or compensated. Otherwise, (2) and (4) lead to the unallowable (3).

If that is so, then there are the following further consequences:

(6) Equality of initial distribution is only available given (1).
(7) Equality of condition is incompatible with self-ownership. For in any redistribution scheme, a given self-owner X would be required to provide service or product by O, which entails (2) above, where Y is the state, and (3) is true of Y.

By ‘initial distribution’ I mean those Noam has already mentioned in his explication of Cohen’s argument concerning the problem of baselines. By ‘equality of condition’ I mean any which implement end-state patterned redistribution principles so that whatever material resources taken as necessary for a given conception of the sorts of beings we are is achieved. If one accepts that (1), as advocated by Nozick, is necessary for any possible initial distribution, allowing for massive inequalities of property and in material goods generally, and thereby allows for massive inequalities of condition further along, then the thesis of self-ownership does come into conflict
with equality. For regardless of whether (6) is in historical fact the natural consequence of (1), given the just steps proviso, (7) is result of (5)—for the attenuating circumstances of justice, such as need, desert, well-being, etc., are achievable by some sort of redistributive process that would entail (2) or (4), which entail the violation of individual liberty that is (3). One cannot argue for the subordinance of (1) to those “attenuating circumstances” understood in the strong sense (the libertarian thesis above) in the service of end-state patterned distributions. 15.

In what follows, I shall here briefly outline some of directions one might take to respond to this challenge. It is crucial to note that the challenge is not just to those these questions. My suggestion is that Cohen’s arguments against equating the thesis of self-ownership with freedom and autonomy, need to be taken seriously considered in answering the question motivating my response, namely: if one accepts that there are certain properly understood facts about the sorts of beings we are, does it follow that individuals have certain obligations, beyond those they voluntarily enter into through contract, to the state and their fellow man?

3. Brief Recap

Thus far in the discussion, Noam has:

i.) presented Nozick’s entitlement theory of justice, focusing on justice in acquisition and rights as side-constraints,
ii.) presented some of G. A. Cohen’s criticisms of Nozick, especially on taking joint-ownership (JO) over common-ownership (CO) as the appropriate baseline situation,
iii.) presented, and argued for, Prof. Scheffler’s argument that the moral basis for Nozick’s theory of rights is more consistent with an alternate, more egalitarian, set of rights;

David has:

i.) focused on the underlying Kantian moral basis for Nozick’s formulation of rights-as-side-constraints, with respect to their form,
ii.) noted that Prof. Scheffler’s alternative set of rights, with respect to their content, is inseparable from the Kantian moral standpoint,
iii.) presented a possible objection that one’s right to their own property is not morally equivalent to one’s right to have their basic needs met, (note
that this sort of objection rests squarely on the sort of needs/entitlement questions I take up in footnote 4, apologies for the two-way traffic between text and notes),

iv.) argued that grounding a theory of justice in any particular moral doctrine is questionable, and explicited how Rawls’ political liberalism avoids those difficulties that might arise from competing moral conceptions of the content of “justice”;

and Adam has:

i.) argued that taking JO over CO as the pre-appropriation baseline offers no improvement, and can itself lead to massive inequalities of condition, ii.) argued also that JO allows for the possibility, even in a two-person situation, that one’s negative rights might be infringed.

What is important to note about the discussion is how the specification of individual rights is   On this view, the fundamental right for libertarians is the right one has to govern themselves as they see fit, which is taken to be derived from a property right they have in themselves. That property right in one’s self is the core of the libertarian position. Keeping the above arguments in mind, I shall now introduce and discuss Cohen’s arguments concerning self-ownership.

4. Distinguishing Autonomy and Freedom from Self-ownership

G.A. Cohen has in a number of articles, and more recently in their collected book-form Self-ownership, Freedom, and Equality (henceforth SOFE, page numbers in brackets) argued that self-ownership neither has the substantive content, nor the argumentative appeal that libertarian thought places on it. In what follows I shall pick up Adam’s objections to Cohen, and then outline some of his arguments against the strong(如果你不理解(强))
in is running out). I shall then conclude by reiterating the upshot of my own discussion thus far: that an redistributive egalitarian program must account for the proper place of (the properly understood notion of) self-ownership within its affirmed set of rights.

If one accepts Noam’s presentation of Cohen’s argument, then it follows that under JO, self-ownership rights are kept intact, and that goods and services can be distributed to those in need. I do not think Adam’s arguments are decisive against choosing (JO) over (CO) as the appropriate “baseline,”
for they do not show how an inegalitarian scramble for raw worldly resources that is possible under (CO) is justified, or even feasible when appropriation of land or resources by one agent, by definition, leaves less for any and all to appropriate (note that Cohen’s argument would not work in a non-scarcity environment, see SOFE, pp.131-3 and p.141) But that of course has not been his point. If I understand Adam correctly, the point is that even a two-person model there are possible problems of collective agreement and resource allocation. If, and it is a big if, the intentional self-starvation of A would cause the death of B, or at least cause B insufferable harm, then it seems reasonable to ask Cohen (SOFE, p.228): where are there limits on permissible harming uses of fully owned objects (in Adam’s example, the use got out of deliberately not using certain resources, while desperately clinging to one’s veto power under joint-ownership) at all, given the possible harming uses, albeit quite extreme uses, of fully self-owned personal powers. However, I do not think that Adam’s objections against choosing (JO) over (CO) as the appropriate baseline dispels Cohen’s point: that that one can allow for the combination of the self-ownership thesis and the egalitarian distribution of resources in such a way that does not allow for the unequal hoarding of resources in the original situation that starts off, and by later iterations of the “just situation by just steps” proviso preserves, a massive inequality of condition as is possible in a situation which maximizes the self-ownership “talent,” i.e. lasseiz faire capitalism.

Are our intuitive notions of freedom and autonomy captured by, and expressed in, the strong principle of self-ownership? Another direction that might prove fruitful is to ask whether the thesis of self-ownership is equivalent to, or the guarantor of, freedom or autonomy. One way of asking the first half of the question would be to simply define the notions in the relevant sense psychological as well as political sense, and compare it. Perhaps asking the second half of the question as it relates to negative rights of non-interference, and moving from a comparison of self-ownership with those rights to explain our intuitive notions of autonomy, would be a more promising start. From that point, one could theorize and formulate what the content of freedom is, and then spell out whatever role self-ownership plays in that formulation. The thought is that our intuitive notions of freedom and autonomy are neither captured, nor in any realistic sense guaranteed, by the thesis of self-ownership, and that might be a useful way in which to begin pursuing the question. Does the unencumbered right to person and property come into conflict with a strong commitment to distributive justice and equality of material condition? Cohen argues it does, for as Noam has presented, it is possible to imagine an alternative baseline that preserves both
self-ownership, and leads to equality of resources. However, the preservation of self-ownership is of a merely formal character Nozick would reject. It is important to note again just what Cohen’s use joint-ownership as the alternative baseline is designed to show: that one can allow for the combination of the self-ownership thesis and the egalitarian distribution of resources (SOFE, pp. 86-91). What is really at issue is whether the self-ownership principle on its own secures, for those do expect the “fruits of their labors,” with the substantive rights to control one’s own lives. If it is considered in the light of the non-interference from taxation that Nozick would like to accrue to the principle, there is much reason to think it does not, and for lack of time, I refer the reader to Ch. 10 of SOFE, pp. 229-244. Placing self-ownership within a set of individual rights, while maintaining a commitment to egalitarian principles, is where the real work needs to be done.

NOTES

1 It is important to note here at the start that the terms “liberal” and “conservative” have flipped-flopped in usage from their advent following the modern democratic revolutions of the 17th and 18th centuries. The emphasis on state power and authority, institutions, obligations between the state and individual in both directions, custom, ritual, and tradition that have usually characterized the “conservative” position have nothing to do with the emphasis on minimal statism, laissez faire free markets, individual liberty, and rights we find in Nozick’s writings, for instance. These values have, especially in the mid-nineteenth century, been the normally conceived as core “liberal” values, whereas now they are more commonly associated with the libertarian position. For more on the origins and meaning of those terms, see R. P. Wolff, The Poverty of Liberalism, and R. Scruton The Meaning of Conservatism.


There are two side issues here that merit bearing out. The first is the relation positive rights create between claims of need and claims of entitlement. If a person has a right to be free from starvation, that is to say that they have a right to food when they would otherwise starve. They need x, and based on that need, they are entitled to x. Hence one’s needs can be used to justify rights to service or products which require extensive state, or collective, resources.

The second issue relates to historical developments that have sharply unraveled a key piece of socialist reasoning. As I say above, these questions are separable, but it is important to note they have not always been so. The standard account of the working class, or that group justifiably deserving the title “proletariat,” was that its members:
1. constituted the majority of society,
2. produced the wealth of society,
3. were the exploited people in society, and
4. were the needy people in society.

Part of the appeal of socialism was that advancing the cause of the working class, so defined, would in broad strokes transform society. The workers had nothing to lose but their chains, because they were in such an inclusive sense those who...In the advanced industrial countries today, (1)—(4) are no longer coextensive. As a matter of empirical fact, (1) and (2) no longer true, except in the trivial sense that all those people who work constitute the majority of society. More interesting for socialist reasoning, and more relevant here, is that (3) and (4) are no longer coextensive with each other. The separation between the needy and the exploited implies either that one has to now choose whether to endorse the Marxist doctrine of exploitation (and the notion of self-ownership embedded within it, see note 15), or to advance the cause of those in need, especially those who do not produce and, as such, are not exploited.


Note, along with David, that this view of rights contrasts starkly with the Rawlsian view, where claims of entitlement are not absolute, but relative to a prior structure of justice that specifies what people are entitled.

Note that whatever one’s conception of “a meaningful life,” there are necessary conditions that constrain it. These include providing those resources
and goods necessary for material subsistence, such as food, which does require broad resources and distributive action by the state. Taking people to be certain sorts of beings with certain sorts of capacities means nothing if they cannot be persons in a substantial sense to begin with.

7 Anarchy, State, and Utopia. p. 169. As Noam pointed out to me, it is crucial to note that consent, and not simply taxation itself, is the real issue. Underlying the claim “taxation being on a par with forced labor” is the belief that people only pay their taxes because of the coercive powers of the state. Nozick would of course be in favor of voluntary aid and other altruistic acts, but in matter of principle, he argues that the welfare of others is no justification for any state to legislate and enforce the redistribution of goods produced by individuals.


10 Anarchy, State, and Utopia. p. 172

11 Of course there are cases of coercion and manipulation that attenuate this remark. However, such processes only point out that force does not govern a given practice we all engage in: the constant choosing of how to conduct ourselves. That choice, and the right to that choice, is what is at issue. Hence the qualification “in some ultimate sense.”

12 Locke, Two Treatises of Government, Ed. P. Laslett. Cambridge: Cambridge University Press, 1960. Sections 27, 37, and 43. Note that these famous statements do not fully bear out Locke’s own views, or wrangling with, the problems with the status of historical acquisition and appropriation. I have deliberately avoided quoting infamous ‘mixing’ passages, for the point is to show Locke’s committal to labor as being an important factor in the creation of value, and that in his account there is some direct connection between the ownership of self, and the ownership of external worldly goods. Note that the second passage is an argument for private property. For more on Locke, see: J. Gough, John Locke’s Political Philosophy. Oxford: Oxford University Press, 1950. and G. Sreenivasan, The Limits of Lockean Rights in Property. New York: Oxford University Press, 1995.

13 Anarchy, State, and Utopia, p. 171

14 see Anarchy, State, and Utopia, p. 161-2

15 There is of course another reason for arguing against the principle of self-ownership: inconsistency on the part of those socialists who condemn
capitalist inequality (that is to say, most all socialists). In the standard Marxist critique of capitalist exploitation, the employer steals from the worker what her work *entitles* her to, namely the benefits that come from the good or service she produced. But that is to say that she has a certain claim on her work as being her own, a claim she could only make with some conception of her being a self-owner, with certain rights to the product of her labor, because *she* produced it.

Now consider the standard Marxist argument against the injustice of capitalism: the distribution of external things, including both raw natural resources and the means by which to produce them, is radically unequal. This unequal distribution has little or nothing to do with considerations of need or welfare, and certainly runs counter to the end-state principle: “From each according to his ability, to each according to his needs.” This inequality is itself the result of initial appropriations which themselves were vastly unequal, and usually in such accounts, acquired unjustly. That inequality of initially unowned resources is taken as matter of empirical fact. However, if one affirms self-ownership in a sense strong enough to formulate an indictment of capitalist exploitation, one has on board the deeply inegalitarian notion that justifies, in part, the ‘initial state’ of inequality taken for granted.

MITCHELL ANDERSON is a fourth year student who is studying Philosophy and Political Economy. He will graduate in the spring of 2004 and then plans on returning to Latin America with his girlfriend where he will write. Mitchell believes that the world cannot be legislated and it is silly to think it can. He also enjoys listening to jazz and asks “Have you heard Coltrane put the sax aside and beat his chest because he had too much music inside him?” That, Mitchell declares is jazz. On Saturday mornings Mitchell drinks coffee and climbs rocks.

ERIN BEEGHLY is a Senior majoring in History with a minor in Philosophy. She will be graduating in the spring of 2004 and plans on attending graduate school. On Saturday mornings, Erin is usually working on estate and antique jewelry appraisals at Lang Antique.

AARON JOSEPH BUTLER graduated from UC Berkeley in 2002 with a degree in Philosophy, however, he has returned to Berkeley to study Chemistry and will receive his Bachelor's degree in 2006. Aaron believes art to be one of the highest expressions of the human mind and enjoys listening to music from many different genres including jazz, blues, rock and world music. On Saturday mornings Aaron enjoys going for a walk in nature.

DAVID JONAS COHN is a junior double majoring in Philosophy and Rhetoric. He is graduating in the spring of 2004 and after that he plans on being a bum, but a well educated bum. This summer he plans on recording some music with Adam Sassoon on the old digital eight track. In his spare time Dave enjoys listening to Radiohead and Bob Dylan and on Saturday mornings, he thinks about what trouble he got into on Friday night.

JULIE ANNE GENTER is a senior who is studying Philosophy, French Literature, and History, in France. After she graduates in May she is going to recover from her studies by traveling, rock climbing, and playing ultimate Frisbee at a few beach tournaments in Europe. She also plans on teaching English in the South of France to support these habits. Julie Anne listens to Radiohead, Elliott Smith, Duke Ellington, Modest Mouse, Charlie Parker, Bach, and David Holmes. Her Saturday mornings are spent making love, buying veggies at the open air market, cooking, and playing Frisbee.
NOAM JACOB is double majoring in Molecular and Cellular Biology and Philosophy and will graduate in the Spring of 2004. After graduating, Noam plans on attending medical school. Currently, Noam enjoys listening to Yoshimi battling it out with the Big Pink Robots. In general Noam likes all music that he thinks is "good." On Saturday mornings Noam goes to Shul.

ALEXANDRA KHREBTUKOVA is a senior double majoring in Philosophy and in Rhetoric, but plans to stay another semester in order to complete her honors thesis in Philosophy. Immediately thereafter, Alexandra will set out on an arduous quest for wisdom so that she may return to her intellectual labors with a mark of authenticity. Alexandra’s musical interests include sessions with a nicely tuned piano and listening to the sounds of a wailing guitar with a thirsty voice behind it. On a Saturday, Alexandra can be found playing hacky-sack by the cathedral near her house, listening to Slightly Stoopid and looking out over the hills of San Francisco.

AVIV ROSENBLATT is well-known for his cover appearances in such well-respected scholarly publications as ‘Seventeen’ and ‘Tigerbeat’. After graduating in May ‘03 he plans to settle in Russia for a while. After that, he plans to resume work with his colleagues at the “National Think Tank for Academic Infiltration.”

PAOLO SANTORIO is a visiting student from Bologna, Italy, where he is a fourth-year student of “Scienze della Comunicazione” (Communication Sciences). Before coming to the U.S. his main interests were semiotics and linguistics, and had never taken a course in philosophy. After his very casual encounter with Analytical Philosophy in Berkeley, Paolo is planning to change his career, and try to come back as a grad student. His favorite musical genre is jazz (and classical, and blues, and…) On a Saturday morning Paolo especially likes to sleep.

KEVIN SALINGER is a Philosophy major who is graduating sometime in 2003. After graduation, he plans on earning a J.D. and a Ph.D. in Philosophy. Kevin listens to Wagner and the Diamanda Galas.
ADAM SASSOON is graduating this May, having majored in Molecular and Cellular Biology with a minor in Philosophy. This summer he plans on spending time with his family (especially his parents and grandparents who he wishes to thank for all of their love and support), writing music, and playing baseball. He would also like to do some traveling if he can swing it. In the fall, Adam will be attending medical school. Adam enjoys listening to a variety of music, but his favorite band is The Beatles. Adam sleeps in on Saturday mornings, and in the afternoons he listens to music and drinks coffee, cursing Dave Shukla’s dreadfully long emails.

TARAK SHAH graduated from Berkeley last December and has not decided on any future plans yet, but is considering graduate school as a possibility. He enjoys listening to Otomo Yoshihide, Noto, Muslimgauze, and The Coup. On Saturday mornings, Tarak drops people off at school and then heads to work because he can’t afford to take the day off.

Dave Shukla completes his course requirements this term. This June he shall either work in a cannery in Dillingham, Alaska, or construction in Long Beach, California, or back in Doe Library. He is currently writing an essay about social explanation.
Last season’s fruit is eaten
And the full fed beast shall kick the empty pail.
   For last year’s words belong to last year’s language
   And next year’s words await another voice.
But, as the passage now presents no hindrance
   To the spirit unappeased and peregrine
   Between two worlds become much like each other,
So I find words I never thought to speak
   In streets I never thought I should revisit
   When I left my body on a distant shore.

~ T.S. Eliot, *The Dry Salvages*