

COGITO

Journal of Philosophy

Vol. IV (New Series) No. 1

2006



Socratic Society of the University of New South Wales
Platonic Society of Macquarie University
Russellian Society of the University of Sydney

COGITO is the Journal of
Socratic Society of the University of New South Wales
Platonic Society of Macquarie University and
Russellian Society of the University of Sydney

Vol. IV (New Series) No. 1, published in Sydney, 2007

© 2006, remains with the authors of the articles

Contributors:

Paolo Bolanos, Macquarie University, PhD candidate
Stephen Bond, Limerick University, PhD
Cressida Gaukroger, Sydney University, Philosophy Honours
Skye Nettleton, Macquarie University, Philosophy Honours, MSGM Masters candidate
Govind Persad, Stanford University, BA, BS
Brad Weslake, Sydney University, PhD
Neralie Wise, Macquarie University, Philosophy Honours

Editors:

Cressida Gaukroger, Sydney University BA Honours, Russellian Society President 2006
Benjamin Schulz, UNSW Masters candidate, Socratic Society President 2006
Mary Walker, Macquarie University PhD candidate, Platonic Society Secretary 2006

Layout, design and artwork:

Gerry Nolan

Cover Art:

Le Bibliophile by Felix Vallotton

Reviewers:

Michaela Baker
Wilson Cooper
Tama Coutts
Russell Downham
Jordi Fernandez
Monte Pemberton
Benjamin Schulz
Robert Sinnerbrink
Mary Walker

Contacts:

Socratic Society:

<http://philosophy.arts.unsw.edu.au/contacts/associations/socraticsoc/socsoc.htm>

Russellian Society:

http://au.geocities.com/russellian_society/

Platonic Society:

<http://au.geocities.com/platonicsociety/>

COGITO

Journal of Philosophy

Vol. IV (New Series) No. 1

2006

Brad Weslake

How Not to Derive Ethical Egoism 5

Skye Nettleton

Loving Existentially: Liaisons Dangereuses and Romantic Love Philosophies of
Simone de Beauvoir and Jean-Paul Sartre 16

Govind Persad

Against Normative Imperialism: Expanding on Dupre 28

Paolo Bolanos

Hegel and the Pathologies of Freedom 37

Cressida Gaukroger

Where Mary went wrong 44

Stephen Bond

Frege and Wittgenstein on Realism 55

Neralie Wise

Is Evolution Progressive? Different Views from two Palaeontologists 62

Cogito is a peer-reviewed journal. The editors are forever grateful to our referees for kindly volunteering to donate some of their precious time toward this project.

Editorial

Cogito started out as the philosophical journal of the Socratic Society of the University of New South Wales. In 2000, the journal became a joint production of the Socratic Society and the Russellian Society of The University of Sydney, and in 2003, the Platonic Society (Macquarie University) came on board. In the current edition of Cogito, we continue this trend toward broadening the scope of the journal. In addition to presenting work from the students of Sydney and Macquarie Universities, we are pleased to include papers from Stephen Bond of Limerick University (Ireland) and Govind Persad from Stanford (USA).

Although Cogito is published by students, and is primarily a student publication, we accept work from across the academic spectrum. In our current edition, we include work from undergraduates, graduates and recently completed PhD students. Older editions of Cogito have also featured essays from academic philosophers, including papers by Isaiah Berlin, David Armstrong and Michael Devitt. We encourage submissions, or proposals for future articles, in the form of essays, reviews, or discussion papers on any matter of philosophical significance, from anyone with a background or interest in philosophical matters. We also strongly urge those who would like to assist with the production of the journal to contact us either by contacting any of the school offices listed below, or via email, at <http://philosophy.arts.unsw.edu.au/contacts/associations/socraticsoc/socsoc.htm> (The Socratic Society), or http://au.geocities.com/russellian_society/ (The Russellian Society), or <http://au.geocities.com/platonicsociety/> (The Platonic Society).

Cogito has been published (somewhat sporadically) since 1966. Copies of most of the previous editions of the New Series - which began in 2000 - remain available for purchase through the Socratic Society. In addition, the previous edition is freely available to download from the Platonic Society's website. Unfortunately, however, the entire back catalogue is not available for purchase. Nor, so far as we are aware, is it even available in storage. This is unfortunate, as the journal gives us a glimpse, albeit a partial one, into the history of philosophy in Sydney. After a concerted effort of inspecting various dusty bookshelves and trawling through a number of dank bookstores, we have managed to recover several older editions, and these have been placed in the UNSW library for safe-keeping. If anyone has any knowledge about the production or availability of older editions, we would greatly appreciate it if they could get in contact with us.

The Editors,

c/o The Socratic Society at UNSW,
or The Russellian Society at Sydney University,
or The Platonic Society at Macquarie University.

How Not To Derive Ethical Egoism

Brad Weslake
University of Sydney

If we view ourselves from a great height, it is frightening to realize how little we know about our species, our purpose and our end, I thought, as we crossed the coastline and flew out over the jelly-green sea.

(W. G. Sebald, *The Rings of Saturn*)¹

Ethical egoism is, it is claimed by a recent commentator (Rachels 2002, 191), a “wicked doctrine that is wickedly hard to refute”. In this paper I aim to provide a refutation of the objectivism of Ayn Rand. The refutation consists of two separate arguments, directed at two separate aspects of the objectivist position. Firstly (in Section I), I argue that the notion of objective function on which the position is based cannot be supported. Secondly (in Section II), I argue that in any case the inference from function to ethical conclusions is invalid. In concluding (in Section III), I make some general remarks concerning the role of evolutionary analyses with regard to function, and offer a diagnosis of a potential source of the problems with the objectivist account.

I

That ethical conclusions might be derivable from claims about functions is a possibility that emerges in *The Republic* (Plato 360BC), where Socrates argues that the function of human minds is what provides the standard against which they are to be morally judged. I will not examine the logic of this argument overall in this paper², instead briefly examining the notion of function on which Socrates relies, by way of leading into a discussion of objectivism. In the brief passage of *The Republic* where Socrates introduces the notion of function (Plato 360 BC, 352d-354b)³, three different senses of the term are used. The following passage introduces two different senses:

‘And would you define the function of a horse, or of anything else, as something one can only do, or does best, with the thing in question?’ (352e)

And soon afterwards Socrates in the course of explaining this definition to Thrasymachus suggests a third sense:

‘But you would do the job best if you used a pruning knife-knife made for the purpose.’ (353a)

The different senses of function used by Socrates in these passages are:

- (1) That which can only be done with the thing in question.
- (2) That which is best done with the thing in question.
- (3) That which the thing in question was designed to do.

Socrates in the course of the discussion does not disambiguate these senses from each

other, and so it is left uncertain whether this is because the senses are supposed to be different ways of understanding the one unified notion, whether each sense is supposed to be sufficient for the purposes of the argument, or whether one sense is supposed to encompass the others. The most natural reading, both in terms of agreeing with the order in which Socrates presents the argument, and in terms of fitting with the overall thrust of the passage, seems to be that Socrates intends (2) as an encompassing definition. On this reading (1) is a trivial case of (2), since if there is only one thing that can fulfil a particular role, it is trivially the best thing for fulfilling that role; and (3) is meant as an explicative device, since something designed for a particular task will in the majority of cases be best suited to fulfil it—this being the very aim of design. No matter which reading Socrates intends, however, each appears on a first analysis as incapable of providing an adequate definition of function. This can be illustrated with the example of the eye, which Socrates seems to use because of the ease with which it will be accepted that it has a function, and the ease with which the conditions for this ascription will be agreed upon. However against (1), it can be argued that an artificial or replacement eye can fulfil the same role—namely, the role of providing sight. Against (2), it can be argued that an artificial or replacement eye might fulfil the role of providing sight better than the eye it replaces⁴. And against (3), we can argue that eyes were not designed, but evolved. This last point requires further examination however, since a widespread view of evolution is that it provides a mechanism for design that does not necessitate a designer (Cummins 1982, 393-394). The issues here have in fact been the basis of a long-running debate about the proper understanding of function in general⁵, and it is a particular view of function in general, based on a particular position within the debate over the status of evolutionary explanations, that forms the ground of the objectivist position⁶.

Before outlining the objectivist argument, it is important to note one feature in particular of the definitions offered by Socrates. Namely, all of the definitions implicitly conceptualise function as being a property of an object; something that an object can possess. That is, depending on which of the senses above is preferred, function turns out to be the property of *that which the object is best at doing*, or *that which can only be done with the object*, or *that which the object was designed to do*. This understanding of function, as a property possessed by an object, I term the *objective* view of function. On the objective view of function, whether an object is functioning well (without qualification) is to be understood as whether the object is functioning well *qua* its objective function. Different views of objective function will provide different accounts of what confers this property on an object, but each will hold that it is this property, and this property only, that provides the standard against which it is to be objectively judged. Turning now to the objectivist variety of ethical egoism, it can be seen that it is just such a formulation of objective function that underpins the argument. The first premise of the objectivist account, as provided by Mack (1971, 736), is that:

If there is some need or requirement which explains (plays a role in explaining) the existence of some thing (object, activity, process, etc.), then that thing functions well if and only if its use or enactment satisfies the need or requirement which explains the existence of that thing⁷.

This is an objective view of function as defined in that it supposes function to be a property of some thing in virtue of facts about the history of that thing—specifically, the “need or requirement which explains the existence of that thing”. And in fact Mack (1971, 737) is explicit that the objective nature of the definition is necessary for the success of the whole objectivist project. Mack does not, however, operate with the same definition of

HOW NOT TO DERIVE EGOISM

objective function that I have provided; instead, he defines objective function as function that is independent of “alteration in persons’ attitude or intentions” (1971, 737). This appears to be a definition that jumps ahead in the argument, however, since Mack has not got to considering persons at all at the point where the definition of function is introduced. The definition I have provided might appear consistent with Mack’s definition, since a property of an object is objective in precisely Mack’s sense—namely, it is not affected by any attitudes or intentions a person may have towards or involving the object. However, the definition I have provided is actually *stronger* than that provided by Mack in that it leaves less room for non-objectivist rival accounts. The reason I have provided a stronger definition of objective function than Mack is because Mack attacks a particular characterisation of function as if it were an attack on all rival, non-objectivist accounts. But I argue in what follows that what Mack attacks is (in terms of my stronger definition) really a kind of relativised form of objective function, and that there remains room for a formulation of an alternative account that does not fall to these objections. And so, I argue, Mack attacks a straw-man. In order to see this, it is necessary to examine Mack’s argument for objective function in more detail.

Mack (1971, 737) identifies a “strong objection” that can be levelled against the premise listed above, in that function might be better defined as the use to which something is put (the “objection” is not so much an objection as a rival account, but I will use Mack’s terms). Mack unhelpfully divides the objection into two parts. What he (1971, 737) terms the *moderate objection* is not worth discussing, since it is hard to see why anyone would take the observation so labelled as an objection at all. The observation is that things may be put to uses other than that which explains their existence. It is quite obvious that this, in itself, is in no way damaging to the objectivist proposal, since it does not even propose a rival account of function, and it is not clear why Mack (1971, 737-739) dedicates almost two pages to iterating the fact. What Mack (1971, 739) however terms the *extreme objection* is closer to the position I aim to defend in this paper—close because not extreme *enough*, at least in the way it is characterised by Mack, as I have suggested. Specifically, Mack notes that the extreme position amounts to the denial that an object has an objective function at all; that it can only be said to have a function in reference to some current goal with reference to which it is being put to use. However as Mack continues, he repeatedly makes remarks of the form “the function of X” (for example at 741) when discussing the extreme position, implying that such (unqualified with respect to some current goal) properties must exist, after all. But of course this is precisely what is being denied. That is, the extreme objection is defined by Mack as the position that the “functions of things are determined solely by their actual uses” (1971, 739). This is true as far as it goes, but misleading, because Mack is even here implicitly using the language of object-and-functional-property that is crucial to the objectivist position (I will not continue to point this out, when referring to Mack, since it continues for the remainder of his paper). What it is important to realise is that this aspect of the objectivist position forms part of what is denied by the extreme objection; it is denied that there exists any such property of an object that is its function. Conversely, all that can be said, according to us extremists, is that a particular object is now being used (more or less effectively) in such and such a way, and that when we talk of the function of an object it is such uses, which are roles in reference to some defined goal, which are being referred to⁸. Mack’s first objection against this position is that it cannot be “strictly maintained” that, for example, we *use* our heart to pump blood (1971, 739; emphasis Mack’s). The point here is opaque; surely we, or at least our bodies, do use our heart precisely for this purpose. Does Mack mean to suggest that because we do not consciously

BRAD WESLAKE

direct our heart to pump we are not entitled to say that we use our heart in such a fashion? It remains the case that this is how the heart is used by the body—surely “use” is not a term that requires an anthropomorphised, conscious *user* (I use my bed to sleep; use requires, simply, a user). Mack himself seems to suggest this when he states that “anything which in functioning satisfies its (*or its users’*) goal is functioning well” (Mack 1971, 737, emphasis mine)⁹. That is, anything which can be construed as goal directed provides a context in which function can be ascribed.

Against Mack (739), the extreme position does *not* imply that a heart may just as effectively have been used for some other purpose than that to which it is currently used; this is a straw-man. To relativise function does not imply that all things have a plurality of available functions, let alone a plurality of functions that they might perform equally well (and for that matter, not even any function that they perform better than any other thing); it is simply to point out that the function of an object is something that can only be defined against a current goal. The objectivist variety of objective function, on the other hand, is crucially *backwards-looking*. That is, the objective function of something is supposed to be provided in virtue of antecedent facts—in particular, antecedent facts regarding the use or need which explains the existence of the object. And so for example when Mack comes to discuss the heart, again (740), with reference to the hypothetical Yogi who is capable of suspending heart activity, the reasoning is as follows. Since it is a requirement for pumping blood that explains the existence of a heart, the objective function of a heart is to pump blood. And therefore, while the heart is providing a useful function for the Yogi (a means for inducing mystical mental states, say), it is nevertheless objectively a bad heart, since it is not performing its objective function. Against this, the extremist argues that two things can here be said about the heart—that it is functioning well for the purposes of the Yogi, and that it is functioning poorly for the purposes of pumping blood. The extremist also says that there is nothing more that needs to be said, and in particular nothing to motivate the introduction of some objective function of hearts that transcends the current uses to which they are being put.

At this point we have two rival views without a basis for discriminating between them (although we might note that the objective position introduces unjustified metaphysical baggage, and might therefore be said to be inferior on grounds of explanatory parsimony), and so further arguments must be provided to prefer one view over another. In order to do this I firstly raise an objection against the backwards-looking nature of the objectivist position, and then defend the extreme position against the arguments made against it (or rather, against a poor imitation) by Mack in the remainder of his paper. The objection against the objectivist position to which I refer was introduced (at least in the particular form here presented) by Davidson (1987), but is perhaps more commonly associated with Millikan (1984). It proposes a hypothetical situation where a lightning strike in a swamp causes a person identical in all physical respects to some other person to be created as it were instantaneously. Such a *doppelgänger* is, in absence of political correctness, referred to in the literature as *Swampman*; and the objection, the *Swampman Objection*. The problem is, straightforwardly, that since Swampman lacks the right kind of history, there can be no ascription of objective function to any of his component parts. We cannot say, for instance, that Swampman’s heart is functioning well, no matter how effectively it is acting as the supplier of nutrients for his life processes, since his heart does not have the right kind of history to have an objective function. Of course according to the relativised position, Swampman’s heart is functioning perfectly well, since it is implicit in this statement that when we refer to the function of hearts without qualification, we are referring to how effectively they are acting as the supplier of nutrients¹⁰. This is considered by many to be a

HOW NOT TO DERIVE EGOISM

knockdown objection against any such backwards-looking account of function—Lalor (1998) calls those of us who share this view *Swampman Intuitionists*. At the very least, backwards-looking accounts require some substantial theoretical manoeuvring to counter the Swampman Objection; and so without delving into this debate here¹¹, we can at this point suggest that there is little reason to prefer the objectivist view of function to the relativised view, unless Mack can provide at least an equally damaging criticism of the latter; and I argue in what follows that he does not.

Before turning to Mack's attempt to critique the relativised view, however, I pause here to flesh out (my own version of) the view in a little more detail. In particular, I want to provide a (very) brief account of how and why we might come to talk about objects *as if* they had objective functions (or at least an outline of how such an account might work) and thus explain why "certain objects do *seem* to have [...] certain objective functions" (Mack 1971, 739-740; emphasis mine). Actually, it is up for grabs whether our everyday talk of function really can be most naturally read as implying the rather philosophical notion of objective function (it doesn't *seem* that way to me, for example); but let's suppose it does. Recall the basic statement of the relativised position—that everything has a function only relative to some purpose. There exist a large class of objects that are used overwhelmingly in a single way, and these are the paradigm objects for discussion with respect to function. Hearts, eyes, hammers, and the like, dominate the literature on function. Because of the singleness of use of these objects, and the fact that they have often been designed (or evolved) to fulfil these precise uses, it is easy to see why the uses to which they are (typically) put are often part of the very concepts of the objects, and that therefore certain uses often become referred to by the same name as the objects that are typically used to perform them. Thus, it is part of the concept of an eye that it is, at least typically, used for seeing; and the act of hitting a nail is perfectly appropriately called hammering, even if it is not performed with a hammer. It can be seen from this observation that if we are not clear about separating the identification of an object and the identification of a function, there exists the opportunity for confusion. Specifically, we might take a statement of the form "Is it a good hammer?" to entail an assessment of the object in itself, when it is really a question about whether it is effective when used for hammering. The question might be asked of an axe, for example, and the answer might well be "Yes, compared to my foot". The confusion stems in part from the ambiguity of such general evaluative questions, in that the standards by which they are assessed vary, depending on their subject matter. That is, the criteria on which I answer the question "Is it a good novel?" and "Is it a good hammer?" are different, the former criteria being relatively abstract and less well defined, while the relative simplicity of the latter criteria is due to the fact that hammers are overwhelmingly used for hammering, and hammering only. With these observations in mind, we are equipped to examine Mack's critique of the extreme position.

Mack (1971, 741) begins by noting what I have just elaborated, that the meaning of "the function of X" is to be interpreted on the extreme position as "the typical purpose or role of Xs". As suggested, this is because this is the mechanism by which the concept of function (at least typically) becomes associated with the concept of a particular object.

But Mack then goes on to criticise this view on the assumption that "the typical purpose or role of Xs" is intended as a *definition* of the function of an X, rather than an *analysis* of the meaning of such talk as "the function of X"; when it is clear that a denial of objective function requires that *no such definition can be provided*. The mistake Mack makes here is to suppose that the extreme view is that objects *acquire* functions in virtue of some typical use of those objects; but this is simply to relativise the objective view, since it retains an idea of functions being properties of objects, against which they can be

objectively judged. Specifically, Mack suggests that the definition cannot be that “having the normal role of an X a necessary condition of being an X”¹², since this would entail that the Yogi’s heart is not a heart; and that the definition cannot be that “satisfying the typical role is a condition for being a good X”, since the typical role of Xs might change without thereby changing what makes for a good X. It is here that the distinction between a definition (which implies the objective view) and a conceptual analysis (which is required by the extreme view) becomes important. This can be illustrated by examining the example Mack provides to support his position.

Mack suggests that if the typical role of medical activity changed from increasing health to increasing sickness, it would remain the case that medical activity *qua* medical activity would retain the function of increasing health. And so in this case medical activity overall would be functioning poorly—not well, as the definition under examination would require. And so the definition fails. But note however the difference, if we take “satisfying the typical role is a condition for being a good X” to be an analysis of how the concept of X incorporates the notion of function, rather than a definition of X; that is, it is to be understood as an analysis of how “medical activity *qua* medical activity” is interpreted as involving a particular function, contextualised against some goal. On this account, it is necessarily part of our current concept of medical activity that it has the function of increasing health. And it is this that provides the standard against which we implicitly judge medical activity. But the issue of how we come to recognise something as being medical activity in the first place is a different matter¹³; and as Mack allows, it may be the case that what we recognise as medical activity comes to (typically) perform a different function than that implicit in the concept of medical activity. In this case, the ambiguity of such general evaluative questions as “is medical activity good” becomes somewhat magnified; or at least it does in the thought experiment, because we seem to be faced with the decision of how to understand the implicit goal, in the absence of any information that might be used to make such a decision. Does medical activity, the concept, in such a world, retain the standard by which we currently understand it? Or does it incorporate a new standard, a standard established by the (now) typical role of what is (still) labelled medical activity? And if it does incorporate such a new standard, don’t we really have a new concept, since we already stated that the function of increasing health is necessarily part of our current concept of medical activity? In such a world, when confronted with the question of whether a particular medical activity is good, we would, if uncertain of the goal against which it was meant, or of the concept being used, simply ask of our enquirer, *what do you mean?* And for current purposes, we can simply observe that the questions raised by the ambiguity are completely irrelevant to the point of the extreme position, which is simply that the questions must be answered and so the ambiguity resolved—that in order to make sense of such evaluative questions *they must be relativised to goals*.

The underlying point here is that we cannot say, without qualification, that any thing is being used badly—that is to say, we cannot make sense of such statements without implicitly relativising them to a particular purpose, and such purposes cannot be said to be either objectively good or objectively bad simply on consideration of the effectiveness with which they are being pursued. As I have argued, Mack fails to provide a substantial critique of this position, due to a mischaracterisation of what such a position entails. Is medical activity in a world where it leads typically to death good? Well, what is the function implied by the concept of medical activity in this world? And is death good, in this world? Both of these latter questions must be answered *first*. We can see from this the way that relativising function makes the ultimate *value* of a particular function a completely separate issue from

HOW NOT TO DERIVE EGOISM

the nature of the function itself, which only provides a standard against which it can be judged effective or ineffective. This prevents the argument for egoism from getting off the ground, because the egoist argument requires that functions can be said to be defective in the sense of “*misdirected*, not merely ineffective” (Mack 1971, 740; emphasis Mack’s).

II

To this point it has been left to one side how it is that objectivism depends on the notion of objective function; that is, the nature of the connection between the objectivist view of function, and the ethical conclusions which are the endpoint of the Randian argument, has not been examined. In this section I argue that the argument supposed to provide this connection is invalid (or perhaps less charitably, *missing*)—and invalid independently of the argument provided above for the failure of objective function¹⁴.

The second premise of the objectivist account provided by Mack (1971, 736), which together with the first premise forms the foundation of the objectivist position, is that:

With respect to each living thing, it is the fact that remaining in existence as a living thing (not merely as a collection of dead cells) requires the successful completion of numerous processes that explains the existence of valuation.

The objectivist reasoning based on this second premise is particularly specious. Without rehearsing the argument in detail, it runs that since the existence of life directly entails valuation, the objective function of valuation is to maintain life. But in terms of the objectivist account of function outlined above, where has the *need* been identified here, that must form the basis of an explanation of the supposed function of valuation? Even if it is true that valuation is a necessary component of a living thing, this does not imply, except in the most trivial sense, that a living thing *needs* valuation. Does a circle need a centre? What has been provided is a piece of conceptual analysis, not the provision of an explanation¹⁵. It is as if it was argued that since being a circle entails having a centre, circles *need* centres. We might tolerate such talk, until we realise that what is being suggested is that we have found an explanation for the existence of centres, and therefore a function for them—to maintain the existence of circles, of course. What we really have here, of course, is the *conceptual* requirement that to be a circle is to have a centre. And to exit the analogy, what we have at most identified with respect to valuation is the *conceptual* requirement that it exist within the context of a living thing. That is, in order to pursue goals, one must be alive. But this is simply irrelevant to the value of goals *per se*, except for the fact that one must remember that achieving most goals is dependent upon the goal of keeping alive. To suggest that this dependency shows that all goals are valuable only insofar as they achieve this subsidiary goal, regardless of anything to do with the goals themselves, strikes one as either misunderstanding the nature of a goal, or being just plain wrong.

The basic error, then, is supposing that a pre-requisite for all ends can be turned into a final end. The crucial move is made by Rand (1969, 16-17) when she attempts to turn a pre-requisite for life into the final goal or end of life; that is, she tries to turn the fact that life is a pre-requisite for values into the fact that life itself is a value. Rand is right when she says that “that which his [man’s] survival requires is set by his nature and is not open to his choice” (22). But as Nozick (1971, 209) observes, supposing death to be a value is entirely consistent with the Randian argument, since life is also a pre-requisite for death.

And of course death is simply a theatrical case; there are potentially infinite goals that require life that may be inserted here. There is no argument in Rand that selects life itself

BRAD WESLAKE

from among these, as a final end (that is, it is not just that there is no *valid* argument; there really is no argument *at all*). Nozick (1971, 209) notes that Rand can only be read as implicitly relying on a premise of the kind “a necessary condition for a value is a value”; but then, as Nozick suggests, presuming that getting cured of cancer is a value would entail that acquiring cancer in the first place also be a value, and this is a patently absurd result. It is a particularly peculiar feature of Rand and her commentators that the reasoning here is not elaborated, even when the deficit is brought to their attention. Thus, in the same way that the crucial move is glossed over in several sentences in Rand’s argument, Den Uyl and Rasmussen (1978) spend more than ten pages reiterating the Randian point that life is necessary for valuation (hardly a contentious or complicated claim), only to make the critical leap in virtually *a single sentence* (1978, 244):

Is there anything that constitutes a final goal or end? From all that has been said [on the fact that life is necessary for valuation], we see that life is the ultimate or final goal of all goal-directed behaviour.

But as Nozick (1971, 209) wryly remarks: “One cannot reach the conclusion that life itself is a value merely by conjoining together many sentences containing the word “value” and “life” or “alive”, and hoping that by some process of association and mixture, this new connection will arise”. The error of Den Uyl and Rasmussen, however, is made all the more curious by the fact that they seem completely oblivious of the fact that the passage quoted above involves the incredible leap of logic under consideration, *even while being aware that Rand has been criticised by Nozick on precisely this point*—since they earlier state, of Nozick’s observation that the Randian argument implies the claim that a necessary condition for a value is a value, that “we do not see how [Nozick can show how this is a premise]” and therefore that they “see no reason to consider [it] at this time” (Den Uyl and Rasmussen, 1978, 241).

III

Before concluding, I briefly examine some ways in which the objectivist account and the relativist account intersect with evolutionary theory, with respect to functions; and offer a diagnosis of a potential source of the problems with the objectivist account. The difference between the objectivist account of function and the relativist account of function (as discussed in Section I) broadly parallels the relationship between the two major theories of function in the philosophy of biology—what Davies (2000, p. 85) terms the theory of selected functions, and the theory of systemic capacity functions¹⁶. The theory of selected functions holds that the attribution of function is equivalent to explaining the existence of the functional item, and that the proper explanation is provided by evolutionary natural selection. As such, it can be seen to parallel the objectivist position. The theory of systemic capacity functions, on the other hand, holds that function is to be understood first in terms of contributions to the behaviour of a wider system. As such, it can be seen to parallel the relativist position. I do not wish here to enter the debate in detail, but simply to point out some further reasons it provides for rejecting the objectivist account.

Firstly, even if the explanation of function can be constrained in terms of the requirement which explains the existence of some thing, there are clear examples where the requirement that provides the proper context of such explanations within living creatures is not the maintenance of life. There are clear counterexamples showing that the maintenance

HOW NOT TO DERIVE EGOISM

of life is a requirement that is neither necessary (for example, proper functioning of the reproductive system in salmon results in the death of the individual salmon, even though selected for by evolution due to reproductive success for the species

overall) nor sufficient (for example, adrenaline secretion reduces fat deposits in humans and so advances life, even though this function of the secretion could not have been selected for by evolution, since obesity was most probably not a significant enough factor in evolutionary selection) for such explanations (McClamrock, 1993). And so while objectivists might want to endorse the theory of selected functions, based on the superficial parallels it exhibits to their position, it does not actually support the result that they wish to establish.

Secondly, and conversely, it is important to realise that the relativist position does not entail a general scepticism about the role of evolutionary considerations in accounts of function. It simply implies that in these analyses function must always be understood in reference to some goal. For instance, to return to the heart, it certainly is the case that it is at least in part due to the fact that pumping blood is useful for the ongoing life processes of an organism that hearts exist. It is important to note two things, however.

Firstly, such a use does not completely explain the existence of hearts, since it is various unrelated factors that play the initial causal role in generating the variation in the first place (McClamrock, 1993)¹⁷, and even then, the benefit of such variation—that which is selected for—must in the first instance be explained in terms of current usefulness to the system, rather than evolutionary history, since initially there *is* no such history (Christensen and Bickhard 2002, 12)¹⁸. Secondly, on pain of repetition, such explanations in no way necessitate an inference from the past use of some thing to some objectively correct future use (adrenaline secretion would have to be *bad* to the extent it reduces fat deposits, on such a reading). This is an inferential step that is completely unrelated to evolutionary accounts. And so while evolutionary accounts can help us to explain hearts and can help to provide a standard context against which hearts may be judged (namely, their usefulness for the ongoing life processes of an organism), they do not establish this context as in itself objectively right or wrong.

What then goes wrong with the objectivist account? Part of the appeal of any functional account of ethics, I suggest, is that it allows for an easy semantic sleight of hand—from talk of functions being good (read here as “effective”) to there being a goodness (read here as “moral goodness”) associated with objects so used. The subtlety of this shift is indicated by the obviousness with which Randians take it to be demonstrated. But to suppose the shift can be made involves a fundamental misunderstanding of ethics. The role of ethics is to discover or construct purposes for us as moral agents—and as I have argued, since functions can only be defined in terms of such purposes, they cannot be used as a basis for deriving ethical conclusions. If to have a function is defined in terms of an external goal, what is the external goal, with respect to humans? What need explains life in general? There is a fundamental error in the presumption that to find things that depend on our existence—or on which our existence depends—is to find the basis for a morality¹⁹.

Written 25 November 2002
Revised 14 September 2003

References

- Ariew, Andrew; Cummins, Robert; Perlman, Mark (Eds.). 2002. *Functions: New Essays in Philosophy of Psychology and Biology*, Oxford: Oxford University Press.
- Branden, Nathaniel. 1970. "Rational Egoism", in *The Personalist*, Vol. 51.
- Christensen, Wayne and Bickhard, Mark. 2002. "The Process Dynamics of Normative Function", in *Monist*, Vol. 85, No. 1, January: 3-28.
- Cummins, Robert. 1982. "Functional Analysis", in Elliot Sober (Ed.), *Conceptual Issues in Evolutionary Biology*, Cambridge MA: MIT Press. 386-407.
- Davidson, Donald. 1987. "Knowing One's Own Mind", in *Proceedings and Addresses of the American Philosophical Association*, Vol. 60: 441-58.
- Davies, Paul Sheldon. 2000. "The Nature of Natural Norms: Why Selected Functions are Systemic Capacity Functions", in *Noûs*, Vol. 34, No. 1, March: 85-107.
- Den Uyl, Douglas and Rasmussen, Douglas. 1978. "Nozick on the Randian Argument", in *The Personalist*, 59: 184-205.
- Hospers, John. 1967. *An Introduction to Philosophical Analysis*, 2nd Ed., London: Routledge.
- Lalor, Brendan. 1998. "Swampman, Etiology, and Content", in *Southern Journal of Philosophy*, Vol. 36: 215-232. Online: <http://philosophy.thereitis.org/pubs/papers.php?path=swamp1.inc> (Accessed 22 November 2002).
- Mack, Eric. 1971. "How to Derive Ethical Egoism", in *The Personalist*, Vol. 52: 735-43.
- McClamrock, Ron. 1993. "Etiology and Functional Analysis", in *Erkenntnis*, Vol. 38, 249-60. Online: <http://www.albany.edu/~ron/papers/funanal.html> (Accessed 22 November 2002).
- Merrill, Ronald E. 1991. *The Ideas of Ayn Rand*. Chicago: Open Court.
- Millikan, Ruth Garrett. 1984. *Language, Thought and Other Biological Categories*, Cambridge MA: MIT Press.
- Nozick, Robert. 1971. "On the Randian Argument", in *The Personalist*, Vol. 52: 282-304.
- Papineau, David. 1996. "Doubtful Intuitions", in *Mind and Language*, Vol. 11: 130-32.
- Plato. 360BC. *The Republic*, Desmond Lee (Trans.), 2nd Ed (Revised), London: Penguin. 1987.
- Rachels, Stuart. 2002. "Nagelian Arguments Against Egoism", in *Australasian Journal of Philosophy*, Vol. 80, No. 2, June: 191-208.
- Rand, Ayn. 1969. "The Objectivist Ethics", in *The Virtue of Selfishness*, New York: New American Library. 13-35.
- Sebald, W. G. 1995. *The Rings of Saturn*, Michael Hulse (Trans.), London: Harvill. 1999.

Footnotes:

- 1 Sebald (1995, 92).
- 2 Briefly, Socrates argues that it is the peculiar excellence of a thing which allows it to perform its function well or badly, and since he has previously argued that justice is the peculiar excellence of the mind, he is able to conclude that the just man will, since exercising the excellence that provides the function of his life, have a good life. The logic here closely resembles that used to formulate the objectivist position, as will be seen as the paper develops.

HOW NOT TO DERIVE EGOISM

- 3 All references are from this translation, and are given in the standard Stephanus notation rather than page number.
- 4 A counter-objection might be made here that when we talk of the function of an eye we are talking of the function of eyes in general; however one can still envisage an artificial eye that provides superior vision than any natural eye; it might, for instance, provide close-range magnification. If it is countered that this is not part of the vision function that the eye aims to fill, then it becomes simply tautological to say that the function of the eye is to provide vision.
- 5 For a recent overview see Ariew *et al* (2002).
- 6 I leave discussion of where the objectivist position sits in relation to these debates until Section III.
- 7 Mack (1971, 737) notes that while such a premise is necessary for the formulation of objectivism, it is left unstated and implicit by both Rand (1969) and Branden (1970). In this essay I concentrate on the formulation of objectivism provided by Mack rather than Rand, since Rand's account is both broader, in that it focuses on giving an overall picture of objectivism rather than a detailed account of its premises, and less precise, in that it is an adaption of a speech and therefore not as clearly organised in terms of an argument.
- 8 This is not to say that it is always the actual current use of an object to which is being referred; it may be the case that the use which best makes sense of talk of an object having a function is some manner in which the object has been used previously, or some typical use for that class of object. I will discuss this in more detail in what follows.
- 9 Rand (1969, 16) makes this explicit when she makes the distinction between "goal-directed" (nonconscious) and "purposive" (conscious).
- 10 It is of course a burden of the extreme view to show how this might plausibly be the case—without appealing to the notion of an objective function, for instance—and in what follows I outline how this can be done.
- 11 Much such manoeuvring has been provided, of course, and it typically involves discounting the value of our intuitions in this area; see Papineau (1996) and Millikan (1984) if your dysfunctional (relative to my purposes) intuitions incline you in that direction.
- 12 Note the switch to the language of assessing an object, here (see Note 13).
- 13 That Mack fails to appreciate this is indicated by his move from talking of the function of Xs to identifying Xs, which I have followed for the purposes of the argument in my treatment here.
- 14 Michael Huemer has pointed out to me that my argument in this section parallels an argument given by Merrill (1991).
- 15 Nozick (1971, 211) emphasises that this is the case by quoting a succinct analysis of the point made by Hospers (1967).
- 16 There are a variety of terms that are used to describe the positions; what is here referred to as the theory of selected functions also goes by the name "the etiological theory", and the functions sometimes by the name of "proper", "selected effects" or "evolutionary" functions; and what is here referred to as systemic capacity functions sometimes go by the name of "analytic" or "causal role" or "Cummins" (after Cummins, 1982) functions (see Davies 2000, 104 n. 1; Christensen and Bickhard 2002).
- 17 McClamrock cites random genetic mutation as *the* factor, though there is reason to believe more is at work in the generation of variation and the operation of selection than just gene mutation (on this, see Christensen and Bickhard 2002, 15 ff.).
- 18 Note the resemblance to the Swampman Objection, which was actually proposed as an objection to etiological accounts of biological function.
- 19 For helpful discussion thanks in particular to Joe Mintoff, and fellow participants in his *Philosophy of Socrates* seminar at The University of Newcastle, 2002.

Loving Existentially: Liaisons dangereuses & romantic love philosophies of Simone de Beauvoir & Jean-Paul Sartre

Skye Nettleton
Macquarie University

Simone de Beauvoir's and Jean-Paul Sartre's life-long experiment for a new *mode d'existence* is a fascinating, inspiring, liberating and outrageously intense love story. In her recent account of Simone de Beauvoir and Jean-Paul Sartre's relationships, *Tête-à-Tête*, Hazel Rowley asks: Why settle for monogamy when you can have "freedom *and* stability, love affairs *and* commitment?" Why keep secrets when you have a best friend to whom you can tell everything? Who amongst us would not wish to frequent the cafés in Paris writing radical, controversial and prize-winning books and spend summers in Italy caught up in the ecstasy of dramatic foursomes?¹

Through a fresh analysis of the lives and philosophies of the lovers Beauvoir and Sartre, making use of dramatic new evidence revealed in Rowley's book, this essay analyses what it means to love existentially.

Simone de Beauvoir and Jean-Paul Sartre met in 1929 and became brilliant young competitors in philosophy. Beauvoir drew second place to Sartre in France's highly competitive teacher's exam, the *aggregation*. It was a hotly-contested jury decision that established Beauvoir's and Sartre's emerging reputations as intellectual equals who would inspire and challenge each other for the rest of their lives. They became highly admired philosophy teachers, writing about and discussing their unorthodox philosophical experiment with love in the cosy, smoky cafes of Paris. They remained at the forefront of France's intellectual avant-garde all their lives. In 1964, Sartre was awarded (but refused on political grounds) the Nobel Prize for his autobiographical narrative *The Words*. Beauvoir had already won France's top literary award, the Prix Goncourt, for her 1954 novel *The Mandarins*.

Sartre discussed all his developing ideas with Beauvoir. He relied on her to pinpoint flaws in his arguments because, Sartre said, she was at the same philosophical level of knowledge, and "she was the only one at my level of knowledge of myself, of what I wanted to do."² He was the first man she considered to be her intellectual superior. He understood and loved her, encouraged and supported her in her work, and wanted to help her become a strong and joyous Valkyrie.³ Rowley describes Beauvoir's feelings about Sartre in glowing terms: "With him she felt extraordinary harmony. There was something incredibly vital about this man. He made her want to discover herself; he made her want to discover the world. With him, she knew she would never stagnate."³¹

As a teenager, Sartre realised he was unattractive and decided it would best to seduce women with his power of speech. He wanted to be "a scholarly Don Juan, slaying women through the power of his golden tongue."⁴ His ardent language and passionate love letters proved to be highly successful in seducing many young women. Plus he was mischievous and loved the game: "I was less keen on the woman than on the play-acting she gave me the opportunity for – since I'd not have agreed to obtain her by just any old means...Possessing her counted for less than the prospects of possession."⁶ He found

the conquests easy, but the game of seduction draining: “I’d come back from a rendezvous, mouth dry, facial muscles tired from too much smiling, voice still dripping with honey and heart full of a disgust to which I was unwilling to pay any attention, and which was masked by satisfaction at having ‘advanced my affairs’.”⁷

Freedom was of utmost importance to Sartre. After seducing a woman, he would insist that she not infringe on his freedom; that she must permit him to sleep with other women because he thought that a great man had to keep himself free. Not wanting to be a hypocrite, Sartre grandly offered the same precious gift of freedom to his girlfriends, saying “it is the most beautiful present I can offer you.” The women were always grateful – or at least they pretended to be. However, says Sartre, “Happily for me...circumstances independent of my will would intervene in time to restore me (after a bit of drubbing) to that dear freedom, which I’d forthwith make haste to bestow upon some other young lady.”⁸

As usual, after falling in love with Beauvoir (affectionately nicknamed “Beaver”), he offered her this gift of her freedom (as if he owned it to give). However, Sartre says, this time, “I was hoist with my own petard. The Beaver accepted that freedom and kept it.”⁹

From this uncertain start they formed a new kind of love relationship, based on overwhelming transparency, strength, Olympian security and happiness.¹⁰ It was an unconventional pact because although Beauvoir and Sartre were devoted intellectual loving companions for life, they did not marry or have children and were not monogamous. They referred to their love as “essential” or primary, but they were free to have other love affairs, which they saw as “contingent” or secondary love relationships.

They felt that jealousy would not be an issue because they promised to be completely honest and open with each other. To prove this, they entered a pact to tell each other every detail of the other relationships and deconstruct every sensation. Sartre called this transparency. “The notion of privacy was a relic of bourgeois hypocrisy. Why keep secrets? As they saw it, their task as intellectuals was to probe beneath the surfaces, plumb the depths of experience, debunk myths, and communicate untarnished truth to their readers.”¹¹

However, they found they could not be completely open because other people were involved so they shamelessly lied to their third-party lovers (and possibly to each other) to spare their feelings and keep them happy.

The physical side of Beauvoir’s and Sartre’s relationship did not last long because, as Sartre puts it, he preferred croissants. Beauvoir explains: “Love was not very successful. Chiefly because [Sartre] does not care much for sexual life. He is a warm, lively man everywhere, but not in bed. I soon felt it, though I had no experience; and little by little it seemed useless, even indecent, to go on being lovers.”¹² When she realised her desires were stronger than his, she embraced the freedom to fall in love with other people.

Sartre and Beauvoir thought that although the individual is “thrown” into existence (i.e. one cannot choose how one enters the world), once a person is conscious, he or she must choose. This is the meaning of “existence precedes essence”. Existential philosophies reject any pre-determined human nature. Humans exist first, starting out as “nothingness” and then define themselves through their choices and actions. We are free to define ourselves and we do so through our actions and projects. We are what we do. This is what Sartre meant when he wrote: “Man is nothing else but what he purposes, he exists only in so far as he realises himself, he is therefore nothing else but the sum of his actions, nothing else but what his life is.”¹³

An existential existence is one which is, above all, free. The individual is free to make choices but is responsible for the consequences of those choices. Because every action is

a choice, we always have choices and are not pre-determined in any way, there is no escaping or, as Sartre put it, “no exit” from our freedom; the individual is “condemned to be free”.

Denying or evading one’s freedom or facticity is, according to Sartre, bad faith (*mauvaise foi*), and is the closest thing to an existential “sin”. One is in bad faith when one allows one’s role or situation to rule one’s choices and actions. Bad faith also arises through self-deception, i.e. lying to oneself by deliberately avoiding facing up to painful facts about ourselves.¹⁴ One might wonder if Beauvoir’s and Sartre’s promise of total honesty to each other was bad faith because such a promise would constrain their freedom; or wonder whether Sartre acted in bad faith by justifying lying to his lovers on the grounds that he was “protecting” them, rather than accepting responsibility for his choices.

Virtue is a concept generally dismissed by existential thinkers. The only thing that comes close to an existential virtue is the personal concept of authenticity, which means being “true to oneself”. Individuals are true to themselves when they engage in projects which are their own and not governed by any externally imposed ideals.

Love is not an abstract concept for existential thinkers. Love exists only in the actions between two people. As Sartre said in his lecture *Existentialism is a Humanism*: “There is no love except that which is constructed, there is no possibility of love except that which is manifested in a loving relationship.”¹⁵

The ideal of an existential romantic love relationship is therefore one in which both partners apply existential principles in their lives and when interacting with each other. They do not act in bad faith or inauthentically because they do not appeal to any role, such as husband or mother, to justify behaviour. Furthermore, they do not appeal to an ideal, such as love, to escape freedom.

Romantic love relationships are among the most complex mysteries of human life. “Romantic love, closely tied to the sensual, has traditionally been seen as something akin to sickness, as a force of irrationality, as potentially anarchic, primitive and insane – the proverbial madman’s disease.”¹⁶

Despite well-meaning intentions and aspiring to lofty heights, love relationships also tend to involve suffering, failure, illusion and disillusionment.¹⁷ Sartre, for example, maintains that: love relationships entail suffering by their very nature; are always conflictual; and are doomed to failure. Beauvoir takes us even further. She suggests that the existence of conflict within love relationships is only half the story and that a genuine equal symbiotic love is possible through emotion.

Sartre was building on Hegel’s idea that objects are what we want more than anything because it is through the recognition that one is different from other objects or people that a person realises he or she is conscious and individual: “Self-consciousness is primarily simple existence for self, self-identity by exclusion of every other from itself...and in this immediacy, in this bare fact of its self-existence, it is individual.”¹⁸

By this Hegel meant that it is via recognition of one’s self through others that one learns what it is to be human; or more specifically, what it is to be a subjective “consciousness” (or freedom). An individual *is* a consciousness, which *is* a freedom because one is able to make choices and is not determined by anything other than one’s choice.

The Other is extremely important because, Sartre says, “The other holds a secret – the secret of what I am.”¹⁹ It is like a buddy system because one requires someone else to interact with in order to understand and define oneself. The Other evaluates one’s actions and gives the actions meaning. Sartre referred to this as “the gaze of the Other.” One

defines oneself through the gaze of another person because the freedom of the self has no way to understand itself except by way of another's freedom.

So in every human interaction, Sartre saw that each person was trying to understand and define his or herself through the Other. Now, in order to define my actions, I need you, the object. But at the same time, you (the object) are trying to do the same to me (the subject). The other wants to make me the object for his subject. According to Hegel and Sartre, you want to possess my subjectivity in order to define yourself. In objectifying me, the Other strips me of my subjectivity. Therefore in engaging with the Other, there is always a struggle over my subjectivity. As such, all relationships are contests or struggles, for which there is never a winner. The Other will always be an object to me. Two people will always be in some degree of conflict because it is always one's subjectivity against another's. "While I attempt to free myself from the hold of the Other, the Other is trying to free himself from mine; while I seek to enslave the Other, the Other seeks to enslave me... Conflict is the original meaning of being-for-others,"²⁰ according to Sartre. As such, "I am a slave to the degree that my being is dependent at the centre of a freedom which is not mine and which is the very condition of my being."²¹ This is why Sartre said "hell is other people"²².

The more highly one regards the Other, the more important the interpretation of one's actions becomes. Sartre says: "The value of the Other's recognition of me depends on the value of my recognition of the Other."²³ Love relations are amplified because when I love someone, I confer upon that person the power of my self-recognition. The problem is that in giving someone the power to define me, "I become hostage to that person's view of me but I do everything I can to control that view."²⁴

According to Sartre, romantic love is merely an attempt to regain control of one's own subjectivity by overthrowing the beloved's freedom.²⁵ It is also a strategy one employs to define oneself in the presence of a person whose freedom one attempts to possess.²⁶ However, ultimately, this is futile because a person can never possess another's freedom.

Because I can never experience or understand the Other's subjectivity, I can never know what he really thinks and therefore, "We are doomed to spend our lives in total individuality."²⁷ Although we try to get close to the Other, especially during sex, we cannot get close enough to understand the Other truly – or to let them understand us truly. While Sartre thought this was hell, Beauvoir took a positive view of the situation.

The Other is still as vitally important to Beauvoir as to Sartre. Through others, one learns to live out one's body, or facticity (everyday life situation), in a certain way. For example, one learns language, intonations and attitudes through others. This is what Beauvoir meant in *The Second Sex* when she wrote "one is not born a woman, one becomes a woman". Sex is biological, but femininity or masculinity is one's gender. Femininity is the cultural overlay that one adds to one's sex as one discovers what it means to be a woman in society.

Sartre does not recognise the debt to others in the same way that Beauvoir does. For Sartre, the Other is always a threat because he is constantly trying to get possession of the subject's freedom. The Other is dangerous and, therefore, there is no obligation to the Other. Beauvoir takes the Kantian approach that there are moral capacities inherent in freedom and that people are aware of the difference between one's will (or passion) and what is 'right'. For Beauvoir, we have a moral duty to others because if one values freedom for oneself, one values freedom for others.

Nevertheless, this is a defective argument because appealing to a moral claim to "love thy neighbour" could be considered bad faith. Moreover, this does not usually follow in

practice because one's freedom comes at the expense of others. However, the important thing for Beauvoir is the simple acknowledgement that we share the world with other people and that we live in society where everyone is to a large extent dependent on the community for survival and self-definition.²⁸

Traditionally, Beauvoir thought, women submit to their lover in order to avoid embracing their transcendence (*pour-soi*). To transcend is to reach beyond one's facticity by actively doing things and constantly changing one's given situation (as opposed to passively accepting one's immanence).

Instead of transcending, the woman in love loses herself in immanence (*en-soi*), meaning that she idolises her lover and gives up her own transcendence for his. "She dreams of transcending her being towards one of these superior beings and of amalgamating herself with the sovereign subject."²⁹ (In Beauvoir's vocabulary, a "sovereign subject" is an independent self-governing human being.)

Instead of asserting herself and defining her own individuality, Beauvoir also says the woman attempts to merge her being with her lover by destroying her individuality: "The woman in love...feels a passionate desire to transcend the limitations of the self and become infinite, thanks to the intervention of another who has access to infinite reality. She abandons herself to love first of all to save herself; but the paradox of idolatrous love is that in trying to save herself, she denies herself utterly in the end. She wants to merge with him, to forget herself in his arms...the desire for a complete destruction of the self, abolishing the boundaries that separate her from her beloved...In order to realise this dream (of ecstatic union), what woman wants in the first place is to serve; for in responding to her lover's demands, a woman will feel that she is necessary; she will be integrated with his existence, she will share his worth, she will be justified."³⁰

Without doubt, there are advantages to being a kept woman: "It is an easy road; on it one avoids the strain involved in undertaking an authentic existence."³¹ Beauvoir explains that such an arrangement can work out very well for a woman: "As one necessary to a being who is absolute necessity, who stands forth in the world seeking goals and who gives her back the world in necessary form, the woman in love acquires in her submission that magnificent possession the absolute. So long as she is loved by and necessary to her loved one, the absolute, she feels herself wholly justified: she knows peace and happiness."³²

This kind of security means that a woman does not have to 'stick her neck out' in the world. She does not risk failure in her ventures because she takes on none of her own. Her man defines the world for her because "the measure of values, the truth of the world are in his consciousness."³³ Not only does she share her lover's prestige and sovereignty in the world, but she shares his identity. She gives up the task of defining herself: "The supreme happiness of the woman in love is to be recognised by the loved man as part of himself; when he says 'we' she is associated and identified with him, she shares his prestige and reigns with him over the rest of the world; she never tires of repeating – even to excess – this delectable 'we'."³⁴

There are three major problems that arise from this state of affairs: bad faith, the risk of sado-masochism and disappointment. The first problem is bad faith because the woman, in loving idolatrously, denies her freedom. Women avoid their freedom and accountability by hiding in romantic love relationships. They may not be happy but they associate the love relationship with success in their life. A woman falls into the trap of bad faith by allowing her world, as Beauvoir describes, to "collapse in contingency, for she really lives in his."³⁵

Bad faith can easily arise for both partners, however: "In loving a woman, what a man

wants is for her to love him. He must present himself to her as one to whom she would want to give her freedom. And she must present herself in such a way that he would want her to want him to want her to love him and so on, seeming to characterise the insecurity of many love affairs.”³⁶

The crisis arises in the attempt to achieve a union with the loved one. “The supreme goal of human love is identification with the loved one.”³⁷ Identifying with a loved one implies an understanding or experience of the partner’s subjectivity. In other words, the object of the game is to contact the Other.

The second problem is that, as with Sartre’s philosophy, there is the risk of sado-masochism because the desire to be loved means affecting another’s freedom.³⁸ A woman in love desires her love to be requited and insists that the lover give himself to her in return. “Her generosity soon becomes exigence,”³⁹ and in generously becoming a slave to her lover, she takes possession of him: “It comes in the form of a gift when it is really a tyranny. Acceptance is in fact an obligation that is binding on the lover, without his having even the benefit of seeming to be a giver; the woman requires him to gratefully accept the burdens with which she crushes him and her tyranny is insatiable...there are no limits to woman’s exigent devotion.”⁴⁰

This is simply another form of the power struggle that Sartre refers to. It is a strategy to gain possession of him, through allowing him to possess her – or at least persuading him that he possesses her. It is an endless game as to who possesses whom.

The third problem, disappointment, arises when the woman discovers that no one is perfect. The man she idolises, the absolute being she gains possession of, who has access to infinite reality, is not a god. He inevitably has faults and he sooner or later becomes “a searing disappointment”⁴¹ to her. Beauvoir says “an authentic love should accept the contingency of the other with all his idiosyncrasies, his limitations and his basic gratuitousness.”⁴² If a woman does not expect too much from a man when she falls in love, she can avoid the disappointment.

Beauvoir’s philosophy of the Other opens the way for a less hostile approach to romantic love relationships than Sartre proposed. While Sartre insists that the gaze of an Other (with whom one chooses to engage with, as opposed to an accidental stare, for example) necessarily involves an ontological struggle for possession of one another’s freedom, Beauvoir believes that in a trusting romantic love relationship, where the lovers recognise each other’s freedom as equal, the possibility for genuine self-discovery without a struggle exists.⁴³ Beauvoir concludes that once the free and equal couple moves beyond the battlefield described by Sartre, a genuine romantic love relationship is possible: “Let both men and women overcome their distrust, and they will find it is possible to restore, in freedom and in equality, the human pair.”⁴⁴

The first step in achieving a free and equal romantic love relationship is for the woman to recognise herself as a free and equal being. The problem from Beauvoir’s perspective in *The Second Sex* was that women in 1949 were not socially free because of their historical situation. While men appropriated the role of subject, women were delegated to the inessential role of ‘the Other’ by subordinating themselves to men. Beauvoir builds on Hegel’s master-slave dialectic but explains that while the slave’s oppression is not voluntary, woman’s is.⁴⁵ As such, “males find in a woman more complicity than the oppressor usually finds in the oppressed.”⁴⁶

For the existentialists, freedom is implicit in one’s consciousness; freedom is ontological. While women have always been *ontologically* free, their psychosocial freedom has been restricted, i.e. the social situation limited their perceived and actual freedom. However, up until recently, it was a rational decision given that the alternatives

and consequences were too costly or unsavoury, such as being ostracised from social circles (though existentialists do not claim to be rational).

Although Beauvoir said that “Everything influences her to let herself be hemmed in, dominated by existences foreign to her own”,⁴⁷ this is bad faith because women voluntarily gave up their freedom and allowed men to dominate them socially.

Women accepted the externally imposed limitations on their social freedom. They existed only in their facticity (their given everyday life situation), without ambition. By losing themselves in immanence, women gave up their own transcendence, their natural right of “becoming”. They defined themselves through their husband, becoming merely an extension of their husband. This existence without transcendence could be defined as “non-being” and is certainly not existential because transcendence is essential to affirming one’s own subjectivity and therefore embracing one’s freedom.

The first step in achieving transcendence, according to Beauvoir, is for women to assert themselves, stand up in the world, “unique and sovereign,”⁴⁸ severing themselves from their dependence on men: “One must first emerge from [the world] into a sovereign solitude if one wants to try to regain a grasp upon it: what woman needs first of all is to undertake, in anguish and pride, her apprenticeship in abandonment and transcendence: that is, in liberty.”⁴⁹

Possibly accepting the context of a capitalist society, Beauvoir wrote that the only way for women to realise transcendence and liberty is through achieving economic independence: “When she is productive, active, she regains her transcendence; in her projects she concretely affirms her status as subject; in connection with the aims she pursues, with the money and the rights she takes possession of, she makes trial of and senses her responsibility.”⁵⁰

With economic independence, Beauvoir contends that women can triumph over their designated subjugation and free themselves from their dependence on men: “It is through gainful employment that woman has traversed most of the difference that separated her from the male; and nothing else can guarantee her liberty in practice.”⁵¹

It must be noted that 50 years ago, there were very strong societal expectations for women to get married and have a family. Women did not originally choose to be mothers; it was their expected duty and role in society. Those who did have their own careers, or chose not to marry, were the exception to the rule.

These expectations exist today, but social circumstances have changed since *The Second Sex*. For example, now women have access to careers through higher education, supported by many years of anti-discrimination and equal opportunity legislation. Women in 2006 can be and often are financially independent. Women no longer need a husband in order to survive as a member of a social community or need children in order to fulfil their societally-imposed vocation as a woman. Marrying and having children, simply because it is socially expected, is bad faith.

Although women are free to choose to live how they please, according to Beauvoir, “there is only one way to employ her liberty authentically, and that is to project it through positive action into human society.”⁵² It is not just about possibilities and intentions. It is about tangible achievements, being active and purposive.

For romantic love to be “genuine, authentic, it must first of all be free.”⁵³ This love requires of a woman that she lives existentially, seizes her liberty and is an independent sovereign person. So armed, she may love in liberty, as a man does, “without putting her very being into question – she must believe herself his equal and be so in concrete fact; she must engage in her enterprises with the same decisiveness.”⁵⁴

Not only is it essential for a woman to be free, she must also be equal to her lover; “one not seeking submission to the other.”⁵⁵ An idyllic heterosexual romantic love relationship can then be achieved as follows: “When a woman finds in the male both desire and respect; if he lusts after her flesh while recognising her freedom, she feels herself to be the essential in the very moment she makes herself object; she remains free in the submission to which she consents...Alterity has no longer a hostile implication; it is in this sense of the union of truly separate bodies that makes the sexual act so moving; it is the more overwhelming as the two beings, who together passionately deny and assert their boundaries, are similar and yet different.”⁵⁶

With equality between partners, the eternal conflict that Sartre describes becomes irrelevant. It is the *mutuality* and respect of the partners that renders a struggle obsolete. “She will have her independent existence and yet continue to exist for him *also*: mutually recognising each other as subject, each will remain for the other an *other*.”⁵⁷

The woman in love has no need to take possession of her lover because he is no longer her salvation. She is actively engaged in asserting herself as an individual and does not need a man to define her existence. Beauvoir says, “Instead of wishing to put man in prison, woman endeavours to escape from one; she no longer seeks to drag him into the realms of immanence but to emerge, herself, into the light of transcendence.”⁵⁸

According to Beauvoir, romantic love, tenderness and sensuality are possible between two free and equal beings: “Genuine love ought to be founded on the mutual recognition of two liberties; the lovers would then experience themselves as both self and as other: neither would give up transcendence, together they would manifest values and aims in the world. For the one and the other, love would be revelation of the self by the gift of the self and enrichment of the world.”⁵⁹

For Sartre, erotic love intensifies the battlefield between the self and the Other because sexual love is one way that lovers try to overcome the hell of each other. While Beauvoir believed that this can be successful and lovers can experience a connectedness – or what she referred to as “intersubjectivity” – through great sex, Sartre thought it impossible. I cannot meet the other as freedom because through my gaze the other becomes an object and his subjectivity escapes me.⁶⁰ Sartre describes the experience as grasping a person who runs away and leaves his or her coat in one’s hands: “It is the outer shell which I possess. I shall never get hold of more than a body.”⁶¹

Because one’s freedom is manifested through the flesh, I make myself flesh to try to lure the Other into manifesting their freedom as flesh. Sartre explains that “In desire, I make myself flesh *in the presence of the Other in order to appropriate* the Other’s flesh.”⁵² However, possession of the Other’s flesh is not the goal of sex. Rather, it is to trap the Other’s freedom within the facticity of his or her body in order to invoke transcendence (the *for-itself*): “Desire is an attitude aiming at enchantment...It is necessary that he be ‘caught’ in it as the cream is caught up by a person skimming milk. So the Other’s For-itself must come to play on the surface of his body; and by touching this body I should finally touch the Other’s free subjectivity. This is the true meaning of possession.”⁶³

Unfortunately, even though one can summon the appearance of the Other’s freedom, it is impossible to possess it, even in bed. “Such is the impossible ideal of desire: to possess the Other’s transcendence as pure transcendence and at the same time as body.”⁶⁴ If possession of the Other’s freedom is the goal of sexual love, the entire rigmarole is pointless because the Other’s freedom is unattainable. And one cannot allow the Other to experience one’s subjectivity either, even if one wanted to. One cannot give away one’s freedom. There is always an irremovable radical separation between people.⁶⁵

Sadism and masochism ensue: “Each one wants the other to love him but does not take into account the fact that to love is to want to be loved and that thus by wanting the other to love him, he only wants the other to want to be loved in turn...hence the lover’s perpetual dissatisfaction.”⁶⁶

Although Beauvoir agrees with Sartre’s “intellectual practical solipsism”⁶⁷ of never being able to know the subjectivity of the Other, she distinguishes between knowing the Other on an intellectual level to knowing or meeting him on a physical and emotional level. This can only happen when there is equality and generosity between the partners: “It is possible to avoid the temptations of sadism and masochism when the two partners recognise each other as equals; if both the man and the woman have a little modesty and some generosity, ideas of victory and defeat are abolished: the act of love becomes a free exchange.”⁶⁸

Beauvoir distinguishes between sex with and without the emotion of love. While accused of preaching sexual promiscuity (and of being nymphomaniac, lesbian, a hundred times aborted, unsatisfied, frigid, repressed, frustrated, among other things)⁶⁹, she is not actually endorsing frequent wanton sex because for her, sex without the emotion of love is a reduced experience.

Through the emotional intoxication of great sex, one can achieve a blissful union with the Other because it is a process of understanding of the Other person as consciousness through the flesh.⁷⁰ The merging is achieved through self-forgetfulness because sex liberates one from one’s own presence. The boundaries of the two individuals are dissolved and they become as if one merged consciousness, thereby creating the possibility of “intersubjectivity”. Beauvoir explains that “Both partners undergo a metamorphosis into flesh through emotional intoxication, and experience themselves and the other simultaneously as subjectivity and passivity.”⁷¹

In *Old Age*, Beauvoir describes sexual love as an adventure. Although the attempt to possess each other’s consciousness through the flesh is doomed, it is not a fight because it is based on mutuality, equality and tenderness. The partners elevate themselves to a level beyond a battle, says Beauvoir: “In the turmoil and desire of sexual activity the consciousness and the body become as one in order to reach the other as a body and in such a way as to enthrall and possess him; there is a twofold reciprocal embodiment. The attempt at possession necessarily fails, since the other remains a subject; but before it reaches its end, the drama of reciprocity is experienced in the act of love in one of its most extreme and most revealing forms. If it takes on the character of a struggle then it begets hostility: more often it implies a ‘togetherness’ that encourages tender affection. In a couple whose love does away with the distance between the ‘I’ and the other, even failure is overcome.”⁷²

Beauvoir’s and Sartre’s erotic love rests on the assumption that our bodies are “manifestations of subjectivity, of freedom.”⁷³ Even though the flesh is transcendence incarnate because the body is the conduit for implementing actions, to bring forth one’s entire subjectivity in a moment of passion is ludicrous. It makes sense that sex is an alternate physical state where two people lose themselves together. However, can a couple’s subjectivity be exposed by reducing them to the level of primal urges and facticity? More likely, it is a metaphor for the other-worldly sensation that one achieves from a heightened sense of one’s own flesh and the closeness of the flesh of the Other. But how an abstraction such as transcendence can be revealed through the flesh (facticity) is difficult to comprehend.

If Sartre was right – that one can never completely understand the Other’s subjectivity

– then it is justified that the objective of a couple in love is to get as close as possible to each other. However, it is still a lost cause because one can never know the Other as subject and even if one could, in a revealing moment, the Other is always “becoming” more than he is and therefore it is instantly lost again. Even if the attempt to understand the other is a lost cause, that does not stop us from trying.

However, perhaps Beauvoir’s focus on emotion refers to something more tangible in sex. Perhaps she was attempting to describe the euphoric experience of sex, where both individuals selfishly satisfy their own desires while amplifying each other’s pleasure at the same time, resembling what one could refer to as a ‘mystic’ or ‘spiritual’ experience. Then, great sex is simply another facet of the symbiotic nature of an existential romantic love relationship: the individual’s freedom to express his or herself comes first, but not at the expense of another’s freedom. Moreover, with an existential approach, the perceived limits of an individual’s pleasure through freedom can be surpassed.

Or perhaps sex is just something that we indulge in to distract our consciousness from reflection on a meaningless and horrible world. Sartre found it in croissants and the seduction of beautiful women. Beauvoir found it in the emotional intoxication of great sex and superlative intimacy.

References

- Bair, D. (1991). *Simone de Beauvoir: A Biography*. London: Vintage
- Barry, J. (1987). *French Lovers: from Heloise and Abelard to Beauvoir and Sartre*. New York: Arbor House
- Beauvoir, S. (1953). *The Marquis de Sade: An Essay by Simone de Beauvoir with selections from his writings chosen by Paul Dinnage*. New York: Grove Press. (“Faut-il brûler Sade?” originally published in *Les Temps Modernes*, December 1951 and January 1952)
- Beauvoir, S. (1968). *Force of Circumstances* (R. Howard, Trans.). Middlesex: Penguin. Originally published in 1963 as *La Force de Chose*. Paris: Gallimard. Retrieved 12/5/2006 from <http://www.marxists.org/reference/subject/ethics/de-beauvoir/1963/interview.htm#n3>
- Beauvoir, S. (1972). *Old Age*. (P. O’Brian, Trans.). London : Deutsch, Weidenfeld and Nicolson
- Beauvoir, S. (1972). *The Second Sex*, (Parshley, Ed. & Trans.) Harmondsworth : Penguin. (*Le Deuxième Sexe* first published 1949)
- Brown, J. (2005, April). The Compelling Nature of Romantic Love: A Psychosocial Perspective. *Psychoanalysis, Culture & Society*, 10, 23-43
- Card, C. (ed.) (2003). *The Cambridge Companion to Simone de Beauvoir*. Cambridge: Cambridge University Press
- Comte-Sponville, A. (2003). *A Short Treatise on the Great Virtues: The Uses of Philosophy in Everyday Life* (Catherine Temerson, Trans.). London: Vintage
- Hegel, G.W.F. (1967). *Phenomenology of Mind* (Baillie, J.B., Trans.). New York: Harper & Row
- Honderich, T. (ed.) (1995). *The Oxford Companion to Philosophy*. Oxford: Oxford University Press
- Rowley, H. (2005). *Tête-à-Tête*. New York: HarperCollins Publishers
- Sartre, Jean-Paul. (1946). *Existentialism is a Humanism* [Lecture]
- Sartre, Jean-Paul. (1956). *No Exit and Three Other Plays*. New York: Vintage Books
- Sartre, Jean-Paul. (1984). *War diaries: notebooks from a phoney war, November 1939-March 1940* (*Les Carnets*). (Hoare, Q., Trans.) New York: Pantheon
- Sartre, Jean-Paul. (1992). *Being and nothingness: an essay on phenomenological ontology* (Barnes, H.E., Trans.). New York : Washington Square Press (Originally Gallimard 1943)

- Solomon, R.C. & Higgins, K.M. (ed.) (1991). *The Philosophy of (Erotic) Love*. Lawrence: University Press of Kansas
- Vintges, K. (1996). *Philosophy as Passion: the thinking of Simone de Beauvoir*. Bloomington: Indiana University Press
- Wagoner, R.E. (1997). *The Meanings of Love: An Introduction to Philosophy of Love*. Westport Connecticut: Praeger

Footnotes:

- 1 Rowley p.xii
- 2 ibid p.335
- 3 ibid p.23
- 4 ibid p.23
- 5 ibid p.78
- 6 Sartre *War Diaries* p.284
- 7 ibid p.285
- 8 ibid p.75
- 9 ibid p.75
- 10 ibid p.274
- 11 Rowley p.xi
- 12 Rowley p.128, Simone de Beauvoir's letter to Nelson Algren Aug 8 1948, *A Transatlantic Love Affair* (New York: New Press, 1998), p.208
- 13 Sartre *Existentialism is a Humanism*
- 14 Honderich p.76
- 15 Rowley p.154
- 16 Brown p.24
- 17 Comte-Sponville p.238
- 18 Hegel p.231
- 19 Sartre *Being and Nothingness* p.475
- 20 ibid p.475
- 21 ibid p.358
- 22 Sartre *No Exit* p.45
- 23 Sartre *Being and Nothingness* p.320
- 24 Wagoner p.96
- 25 ibid p.96
- 26 Solomon & Higgins p.227
- 27 Vintges p.52
- 28 ibid p.54
- 29 Beauvoir *The Second Sex* p.653
- 30 ibid p.660
- 31 Rowley p.195
- 32 Beauvoir *The Second Sex* p.663
- 33 ibid p.662
- 34 ibid p.662
- 35 ibid p.662
- 36 Wagoner p.104
- 37 Beauvoir *The Second Sex* p.662
- 38 Barry p.299 on Sartre *WarDiaries* pp.255-61
- 39 Beauvoir *The Second Sex* p.666

LOVING EXISTENTIALLY

- 40 *ibid* p.666
41 *ibid* p.664
42 *ibid* p.664
43 *ibid* p.17
44 Bair pp.415-6
45 Card p.152
46 Beauvoir *The Second Sex* p.731
47 *ibid* p.721
48 *ibid* p.721
49 *ibid* p.720
50 *ibid* p.690
51 *ibid* p.690
52 *ibid* p.687
53 *ibid* p.491
54 *ibid* p.705
55 Bair pp.415-6, referring to the article "It's About Time" written for the American magazine *Flair*, in which Simone describes her idea of the "new face" of love (*Flair*, vol. 1 No. 3, April 1950 pp.76-77)
56 Beauvoir *The Second Sex* p.422
57 *ibid* p.740
58 *ibid* p.726
59 *ibid* p.677
60 Vintges p.49
61 Sartre *Being & Nothingness* p.511
62 *ibid* p.506
63 *ibid* pp.511-2
64 *ibid* p.512
65 Vintges p.54 on Beauvoir *Pyrrhus et Cineas* p.330
66 Rowley p.100
67 Vintges p.62
68 Beauvoir *The Second Sex* p.701
69 Beauvoir *Force of Circumstance* pp.196-200
70 Beauvoir *The Marquis de Sade* p.33
71 *ibid* pp.32-33
72 Beauvoir *Old Age* pp.318-9
73 Wagoner pp.98-99

Against Normative Imperialism: Expanding on Dupre

Govind Persad
Stanford University

1. Introduction

The natural and social sciences frequently aspire to an ideal that portrays intradisciplinary biases and commitments, as well as those of the researcher, as nonexistent or unimportant. This ideal of objectivity suggests the further notion that scientific theories are normatively neutral: that a given scientific theory, rather than implying or suggesting a particular view about values, is compatible with any normative position. A second commonly held scientific tenet is that the “horizontal expansion of theories is an unqualified scientific good”:¹ whether or not chemists, for example, believe that the chemical sciences can be vertically reduced to particle physics, they believe that interactions at a given level of structure (for example, atoms and molecules) can be united under the rubric of chemistry.²

This move towards a horizontally united model of science, in which the study of each level of structure is unified under a single scientific methodology, is not without its critics: John Dupre, in his “Against Scientific Imperialism,” asserts that “[t]he most pressing questions about horizontal unification arise in the study of human behavior,”³ and argues that these questions, at least within the domain of behavior, call the horizontal unification of science into doubt. Dupre sees both economics and biology as attempting to unify the study of human behavior under their own respective banners, an attitude which he terms “imperialist”:

[A]t least two approaches to human behavior, those deriving from economics and evolutionary biology, often involve clearly imperialist tendencies. Devotees of these approaches are inclined to claim that they are in possession not just of one useful perspective on human behavior, but of the key that will open doors to the understanding of ever wider areas of human behavior.⁴

The “scientific imperialism” against which Dupre inveighs here just is the placing of one’s own scientific discipline as the agent and beneficiary of a new unified science. Dupre then investigates and critiques the claims made by these theories to provide a unified account of human behavior, and concludes by arguing against horizontal unification. I concur with Dupre, but suggest in addition that the case of human behavior not only challenges the claims of economics and biology to horizontal unity, but also undermines their claims of normative neutrality. I will consider three aspects of the economic and biological explanations of behavior that violate normative neutrality. My first example will be the normative commitments to self-interest, both covert and overt, present in economic and biological theories of behavior. In doing this, I will also consider the self-referential nature of both economics and biology. Both disciplines are self-referential in that they place the supposedly neutral and removed researcher within her own domain of study. By simultaneously encouraging researchers to see themselves as economic or biological

beings and positioning themselves as sciences able to explain the reasoning of such beings, they circumscribe researchers' ability to critically reason about their disciplinary commitments. Finally, I will highlight the nonfalsifiability of biological and economic accounts of human behavior, due to their shared propensity to provide a disciplinary story to explain all results, even apparently discrepant ones. Taken together, these examples illustrate an overarching stance of "normative imperialism," under which normative possibilities are either dismissed entirely or arrogated exclusively to the proposed scientific theory, whether economic or biological.

Ultimately, I do not argue that economics and biology should be excluded from an explanation of human behavior, but rather contend, as Dupre does, that these disciplines should instead comprise important but not exclusive parts of a "horizontal pluralism" of theories about human behavior.⁵ Dupre suggests that, when human behavior is the phenomenon under study, theories such as "psychology... anthropology, [and] sociology" also deserve a place in a horizontal pluralism of science.⁶ When human behavior's normative dimensions are under consideration, I argue that this pluralism should expand to also include theories, such as the traditional humanities (e.g. history, philosophy, and literature), that recognize the human capacity for creativity and moral commitment.

2. Normativity and Self-Reference in Economics and Biology

Dupre highlights the normative position implicit in much of economics when he discusses the commitment of rational choice theory to an implicit normative stance that valorizes self-interest: "whereas in principle the rational choice perspective may be capable of encompassing any mix of self-interested and non-self-interested concerns, in practice it almost invariably reflects the assumption that the former are far more significant."⁷ To support this point, he notes that empirical research has illuminated hidden normative commitments connected with economics, and cites the work of Robert Frank as an example of this concealed normative position:

Frank administered tests of the disposition to cooperate to students before and after completing introductory courses in microeconomics and in astronomy, and found that the economics students were strikingly less cooperative at the end of the course.⁸

Frank's result illustrates that the study of economics has altered the behavior of economics students, at least within the economic domain. This alteration causes the students' behavior to conform more closely to the views about cooperation that economic theory hypothesizes. Since the change in behavior results from a change in beliefs about value that is caused by learning economic theory, economic study (at least as taught in the course Frank uses as an example) can be characterized as having implicit normative commitments.

This example suggests that implicit normative commitment can result from theoretical self-reference. Instead of merely being a descriptive theory of what people *in fact* do, economics is here, knowingly or unknowingly, producing a prescriptive position on what people *ought* to do. The economic paradigm position, that people are motivated by self-interest, thus does not become true because it accurately describes objective facts about human behavior that precede theory, but rather because it alters the world in order to produce these facts.

In a later paper, Frank also suggests, both on theoretical and empirical grounds, that education in behavioral biology may create results similar to those caused by economic education:

On theoretical grounds, we would not have included biology students as a control group because the theoretical paradigm in behavioral biology stresses self-interest perhaps even more strongly than does the corresponding paradigm in economics. Thus, we note with interest that the movement towards dishonesty was larger for biologists than for others.⁹

Thus, just as learning economics may make the economics student more of a self-interested *homo economicus*, learning biology seems to alter the beginning biologist's behavior in a way that confirms the biological paradigm. Had Dupre been aware of this later result, he might well have found this similarity between biology and economics particularly interesting because of the connections he draws between economics and biology. Both, according to Dupre, are disciplines which investigate human behavior in an imperialist way, and both, as shown by Frank, have a self-confirming and normatively committed character.

As has been shown, this hidden normative commitment in both theories seems to at least partly stem from the self-referential position in which economists' and biologists' own theories place them. It may also result from the model-based rather than law-based nature of economic and biological theories. As Dupre suggests,

Advocates of the semantic view of theories have emphasized that much of science is more realistically described as consisting in sets of models than in universal laws. It seems to me that this is indeed a more realistic account of the way science is practiced in biology and economics.¹⁰

The result of a model-based conception of science is a pressure to conform to the model in question. Frank references this idea of economics as a predictive model in his own later paper: “[I]t would be remarkable indeed if none of the observed differences in behavior were the result of repeated and intensive exposure to a model whose unequivocal prediction is that people will defect whenever self-interest dictates.”¹¹ As Frank states, “exposure to the self-interest model does in fact encourage self-interested behavior.”¹² While Frank makes this claim in reference to economics, this comment would be equally relevant to biology, since it also conceives of human behavior in self-interested terms.

This result that exposure to a view can influence the behavior of practitioners, an aspect of both economics and biology, may require a model-based conception of science as well as self-reference. Physics students are subjects of the laws of physics, and so seem to fall under the same self-referential picture described above. However, it would be surprising—and indeed almost impossible—if students who completed introductory courses in physics were found to behave more in accordance with the laws of physics than those who completed coursework in English or political science do. This is because physical laws (unlike legal edicts, for example) are descriptive laws: they do not assert what *should* happen, but describe what *does* happen. While physical laws generally represent simplifications or idealizations of real phenomena, they generally attempt to describe approximately what happens in the world. In contrast, the stances taken by economics and biology have a normative flavor: they are prescriptive, rather than descriptive.

This difference in effects between prescriptive versus descriptive law also produces a very different attitude towards violations of the two different types of law. A violation of a descriptive physical law is taken to be a counterargument against the *theory*, while a violation of a prescriptive law, such as a legal stricture, generally leads to the *violator* being censured. When people act in a way contradictory to self-interest and thus to what economic and biological laws predict – for example, when they cooperate in a Prisoner's

Dilemma or help non-kin who will not reciprocate – they are frequently criticized (at least within the theory) for being irrational, which amounts to not conforming to what the model predicts. This pressure to conform to the model can produce a lack of willingness to consider normative alternatives to theoretical predictions, for fear of turning out to be “soft” or a “sucker” who cooperates when theory predicts defection: “By emphasizing the prevalence of opportunism, economics training may lead students not to engage.”¹³ This disengagement can in fact lead to worse results than cooperation.¹⁴

Thus, economic and biological theories of human behavior, due to their self-confirming and model-based character, contain preexisting normative commitments to the theoretical conclusions that they attempt to derive from scientific analysis. Both theories centrally commit themselves to the idea that self-interest is the primary motivator of human behavior, and thus both tend to produce self-interested behavior in the world. Though intuitively problematic, this disciplinary feature is not necessarily an irreparable problem for economics and biology: however, the normative pull of the theory must be admitted and taken into account.

3. Falsifiability and The Limits of Scientific Explanation

Before I consider the ways in which this normative commitment to self-interest can be dealt with, I will first discuss a second commitment, to a method of explanation rather than a conclusion. Not only do economics and biology tend to produce self-interested behavior, they also tend to discover these results when observing this world. Economists tend to see human behavior as exemplifying an effort to maximize utility, while behavioral biologists tend to see human behavior as reflecting evolutionary adaptation. Since a trait is adaptive when it increases an organism’s own reproductive fitness, this explanation also couches behavior in self-interested terms.

The above commitment to a certain type of explanation raises the worry that economics and biology, at least in reference to human behavior, fail to make their theoretical claims falsifiable. Karl Popper suggests that one feature of a scientific theory is that it ought to be falsifiable by evidence: that is, it should be possible that some facts, if they were true, should count as evidence against the theory.¹⁵ This sort of argument is often used to reject arguments that God created the world, for example: since it is not possible to provide experimental evidence that would refute the creationist claim, the claim is not testable.

This falsifiability criterion presents a concern for some pictures of rational choice theory in economics. As Dupre states, the economic definition of utility is “vacuous,”¹⁶ because it expands to fit any possible behavior:

If it is true that people are motivated by some concern for the well-being of others, or some wish to behave in ways customary among their social group, or according to their conception of morality, duty, etc., then behavior satisfying these various criteria will simply be seen as providing some contribution to the utility of the agent.¹⁷

This definition of utility does not allow for any behavior to be performed for reasons other than contribution to utility, since utility is just whatever a person most prefers. Thus, the economic theory of utility is, according to many, not falsifiable, since a utility-based story about why a behavior happens can always be told.¹⁸

Falsifiability, or the lack thereof, is likewise a concern with the model of natural selection as determining traits. This concern is stressed by Lewontin and Gould, who argue that adaptationist reasoning follows a “Panglossian paradigm” in which a story can be told that explains any behavior or physical feature of an organism in terms of improved selective

fitness.¹⁹ This method of argument seems to violate the falsifiability criterion because this general approach is not subject to disproof by counterexample, since “the rejection of one adaptationist story generally leads to its replacement by another”²⁰ rather than the abandonment of the adaptationist model as the appropriate explanation for the behavior studied. Just as any behavior can be explained as utility-maximizing, it seems that any behavior can be explained as adaptive.

How can Popper’s falsifiability criterion be met? Dupre hopefully suggests that some biological stories will be ruled out as implausible:

Although evolutionary biologists have notoriously been accused of making up any old story that will link some observed behavior to reproductive success, the need for such a story is a non-trivial requirement, and stories can be investigated for plausibility.²¹

Likewise, Sen suggests, contrary to Dupre’s claims about rational choice, that economic models of utility could be both falsifiable and false, if it proves the case that utility is not a plausible explanation of a given behavior.²² However, because of the wide range of potential stories and the low threshold for plausibility, adaptationist and utility-based explanations seem perpetually available regardless of attempts at disproof. Lewontin gestures at this problem:

[W]hen I once pointed out to a sociobiologist that sane and rational human beings were willing to go to prison rather than engage in armed struggle, he replied that their resistance to the state was a form of aggression. One need not be an orthodox follower of Karl Popper to see that a theory that allows things to appear in the form of their apparent opposites when convenient is not of much value.²³

This level of commitment to an adaptive explanation, to the exclusion of clearly apparent contradictory evidence, violates the criterion of falsifiability.

Dupre suggests that one result of the push for a certain sort of theoretical explanation is that modeling human behavior by either method involves a strong push for theoretical self-confirmation:

I believe that the inadequacies of applications of such models so far from their natural home would seem obvious in most cases if there were not some strong motive inclining scientists to be sympathetic with the projects they represent. And indeed the standard response to such skepticism is to admit that the treatments are as yet too simplified to be altogether realistic, but that they constitute the first step towards some very important goal. What goal? The goal of applying the preferred imperialist strategy of explanation to the phenomena in question.²⁴

Here, we again have a self-referential model, this time favoring a theoretically prescribed method of explanation rather than a theoretically predicted conclusion. Scientists are not objective, but have an interest in the success of their own model or explanatory strategy. As such, rather than being willing to reject an explanatory strategy, they persist in telling alternative stories when the data do not fit the desired outcome. This privileging of their own model leads to the persistence of an economic or adaptationist explanation of behavior even in the face of other plausible alternatives.

Like the normative commitment to self-interest, a commitment to providing explanations in terms of utility or adaptive function can have real consequences. A person who sees the world in terms of pure utility would be equivalent to Sen’s “rational fool,” who

defines his entire life in terms of a preference ordering, without differentiating between his welfare, preferences, and commitments.²⁵ This flattened picture of motivation tends to dismiss the difference between moral commitments and desires or preferences, a stance Dupre criticizes: “to treat altruism, morality, or accepted social norms simply as tastes that some people happen to have—I like candy and fast cars, you like morality and oysters—is grossly to misplace the importance of norms of behavior in peoples’ lives.”²⁶ A person who is committed to an adaptive explanation of behavior likewise encounters difficulties in arguing against behaviors that are considered reprehensible: consider the justified furor over Thornhill and Palmer’s *A Natural History of Rape*, which argues that rape results from natural imperatives.²⁷

4. Responding To Normativity

There are two stances the economic or biological theorist can take towards the normative results detailed above. The first is to explicitly embrace the apparent normative results of the theory and take up the position of arguing for a particular ethical stance. Garrett Hardin adopts this approach in his essay on the commons problem, where he asserts that standard ethical arguments use exploitative means such as guilt, and argues instead that ethical arguments ought instead to appeal to self-interest in order to be both successful and morally permissible.²⁸

This move, however, places scientific theory’s claim to neutral objectivity in question. One approach is to give up the privileged position of science and reposition the theory as a normative one. James Alvey notes that economics was originally a branch of moral philosophy, and continued to be considered as such during the time of Adam Smith.²⁹ Only later, with the abandonment of inquiry into individual preferences and the positioning of economics as akin to natural science, was this normative stance abandoned.³⁰

The second approach is to attempt to retain scientific objectivity while taking an (often covert) normative position: often, this is allied with the assertion that biological or economic facts imply normative claims. This sort of claim is asserted by Yezer, Goldfarb, and Poppen with reference to economic education:

If students in economics courses learn that the world is in actuality less cooperative than they initially and incorrectly perceived it to be, then the teaching that produces this result should be viewed in a positive light.³¹

Thus, they assert that learning about facts in the world is good, regardless of the normative results of this education. There are two problems with this claim, however. First, the “facts in the world” that Yezer et al. reference may well be created by the economic theories they endorse, rather than being pretheoretically present. This suggests an interesting equivocation at the heart of economic and biological laws: they claim both the inherent authority of a natural law and the normative force of a prescriptive law, but having one seems necessarily to block the other. Second, the fact that an “ought” does not follow from an “is” is a well-known logical dictum and presents a problem for any attempt to divine the appropriate normative stance from empirical evidence.³² The students may learn a fact about the world, but the appropriate response may well be to change that fact.

Alternatively, rather than taking a normative stance, economists and biologists can attempt to retain scientific neutrality or at least scientific balance by becoming aware of the concealed normative biases and/or providing counterbalancing normative views. This is what is suggested implicitly by Frank, Gilovich, and Regan in their example of an introductory microeconomics course that does not greatly decrease cooperative behavior;

this course, instructively, was not taught by a rational choice theorist, but by a specialist on economic development in China.³³ They expand further on this idea by suggesting that “with an eye to both the social good and the well-being of their own students, economists may want to stress a broader view of human behavior in their own teaching.”³⁴ This is not the Sidgwickesque suggestion of concealing the idea that self-interested immorality can be an effective approach to life,³⁵ but rather an attempt to provide a fair rather than normatively biased picture of human behavior.

There may still be some places where it makes sense to draw normative conclusions or make normative decisions based on scientific results: for example, discovery of a new economic inequality or biological frailty might motivate further investigation. Likewise, the fact that a known economic or biological state (such as unemployment or illness) is seen as undesirable might legitimately lead to research into these conditions. The normative basis for the judgment of undesirability, however, lies outside the biological or economic theory, rather than within it. The point of considering normative implications is not to create an unbreachable wall between normativity and objective science, but to appropriately address normative issues when they intersect with scientific study.

5. Conclusion: Normative Pluralism and Human Behavior

How can one do economics or biology in light of the potential pitfalls on offer? Dawkins offers one potential picture, counseling us that, even though biology suggests a model to us, we are free to provide actual counterexamples against it when it applies to human behavior:

[Dawkins] suggests that conscious foresight may enable us to develop radically new kinds of behaviour. “We are built as gene-machines and cultured as meme-machines, but we have the power to turn against our creators. We, alone on earth, can rebel against the tyranny of the selfish replicators.”³⁶

While this suggestion may not work perfectly, it merits consideration. Likewise, an ethically conscious approach to economics would suggest that even though economic theory predicts noncooperation if we behave exclusively as *homo economici*, we may be able to resist this theoretically predicted pull once we know it exists. The best normative use of scientific theories may ultimately be to help us realize our temptations: if we know that economics and biological models predict that we may act in a way that does not further our actual aims, we can use that knowledge to recognize when we are being diverted from our normative aims by biological or economic pressures, and act accordingly. The self-knowledge provided by understanding biology or economics can thus be an aid rather than a handicap in carrying out our normative projects.

Where can normative knowledge come from? Here is where I suggest that Dupre’s essay falls somewhat short. The initial normative values against which we can analyze the influence of scientific theory come not necessarily from the disciplines Dupre lists (such as psychology and sociology), but also from the traditionally “humanistic” disciplines such as history and philosophy. Such education can be combined with traditionally scientific models to provide a greater overall picture of the discipline: as Alvey asserts about the teaching of economics, “One way of restoring a place for ethics in the curriculum is to include a segment on Adam Smith’s *Theory of Moral Sentiments*.”³⁷ Biologists studying human behavior could likewise learn from Darwin’s *The Descent of Man* or Gould’s *The Mismeasure of Man*.

This response coheres with Dupre’s complaint about the mathematical abstraction of the economic and biological models. Dupre criticizes Gary Becker’s *Treatise on the Family* as marred by a “continuous and obfuscating veneer of mathematics,” and suggests that an “intelligible exposition of the ideas in question, unobscured by jargon or irrelevant formalism” might suffice to undermine abstract economic models such as Becker’s market-based account of marriage and the family.³⁸ Commenting on Becker’s account of marriage, for example, Dupre notes:

To mention only one striking point pertinent to marriage “markets”, the conception of romantic love which, for all its ideological problems, surely plays some role in matrimonial decisions in Western societies, is precisely an extreme commitment to the non-substitutability of one item for another.³⁹

I suggest that the concept of love and its impact on human behavior might better be understood by reading Plato’s *Symposium* or Tolstoy’s *Anna Karenina* than by attempting to quantify it in economic or biological terms. Lewontin concurs with this attitude in his rejection of the attempt to model culture in evolutionary terms: “We would be much more likely to reach a correct theory of cultural change if the attempt to understand the history of human institutions on the cheap, by making analogies with organic evolution, were abandoned.”⁴⁰ In place of this evolutionary analogy, Lewontin instead calls for a deeper analysis of the history of ideas.

Lewontin and Gould, at the end of their essay, call for a reintroduction of the unreduced, whole organism into biology via a pluralist theory: “A pluralistic view could put organisms, with all their recalcitrant yet intelligible complexity, back into evolutionary theory.”⁴¹ The pressure to reduce behavior to models too often acts to conceal the true complexity of human behavior: in contrast, I believe that a pluralism of different theories would better be able to deal with, rather than avoid, complexity in real-world problems. De Waal suggests as much when he criticizes the exclusive biologizing of the social problem of rape: “[I]t would have been more productive to join forces and consider a wide range of perspectives on an ugly behavior that has harmed so many.”⁴² Ultimately, I argue that the human being – the *humanistic* human being, with norms, values, and commitments – deserves to be genuinely considered, rather than merely explained away, and that this is why we need a pluralist rather than imperialist approach to human behavior.

Footnotes:

- 1 John Dupre, “Against Scientific Imperialism,” *Philosophy of Science Association Proceedings* 2 (1994): 374.
- 2 Ibid.
- 3 Ibid.
- 4 Ibid.
- 5 Dupre, 377.
- 6 Dupre, 377.
- 7 Dupre, 378.
- 8 Dupre, *ibid.*
- 9 Robert H. Frank, Thomas Gilovich, and Dennis T. Regan, “Do Economists Make Bad Citizens?,” *Journal of Economic Perspectives* 10.1 (1996): 189
- 10 Dupre, 380.
- 11 Frank, Gilovich, and Regan 1993, 170.

- 12 Robert H. Frank, Thomas Gilovich, and Dennis T. Regan, "Does Studying Economics Inhibit Cooperation?" Journal of Economic Perspectives 7.2 (1993): 59.
- 13 Frank, Gilovich, and Regan 1996, 191
- 14 Frank, Gilovich, and Regan 1993, 170
- 15 Stephen Thornton, "Karl Popper," Stanford Encyclopedia of Philosophy (Fall 2001), ed. Edward N. Zalta.
<<http://plato.stanford.edu/archives/sum2005/entries/popper/>>
- 16 Dupre, 377.
- 17 Dupre, 377.
- 18 Amartya K. Sen, "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory," Philosophy and Public Affairs 6.4 (1977): 325.
- 19 Stephen Jay Gould and Richard Lewontin, "The Spandrels of San Marco: A Critique of the Panglossian Paradigm," Proceedings of the Royal Society 205 (1979): 587.
- 20 Ibid.
- 21 Dupre, 379.
- 22 Sen, 325.
- 23 Richard Lewontin, "The Wars Over Evolution," New York Review of Books 53.16 (2005).
<<http://www.nybooks.com/articles/18363>>
- 24 Dupre, 378.
- 25 Sen, 336.
- 26 Dupre, 380.
- 27 Frans DeWaal, "Survival of the Rapist," New York Times, April 2, 2000 <<http://www.nytimes.com/books/00/04/02/reviews/000402.002waalt.html>>
- 28 Garrett Hardin, "The Tragedy of the Commons," Science 162.3859 (1968): 1247.
- 29 James E. Alvey, "A Short History of Economics As a Moral Science," Journal of Markets and Morality 2.1 (1999): 1234-37.
- 30 Ibid.
- 31 Anthony Yezer, Robert Goldfarb, and Paul Poppen, "Does Studying Economics Discourage Cooperation? Watch What We Do, Not What We Say or How We Play," Journal of Economic Perspectives 10.1 (1996): 178.
- 32 John Mackie, "The Law of the Jungle: Moral Alternatives and Principles of Evolution," Philosophy 53 (1978): 747.
- 33 Frank, Gilovich, and Regan 1993, 168.
- 34 Ibid., 171.
- 35 Barton Schultz, "Henry Sidgwick," Stanford Encyclopedia of Philosophy (Fall 2001), ed. Edward N. Zalta.
<<http://plato.stanford.edu/archives/win2004/entries/sidgwick/>>
- 36 Dawkins, qtd. in Mackie, "The Law of the Jungle", 747.
- 37 Alvey, 1247.
- 38 Dupre, 378-79.
- 39 Dupre, 380.
- 40 Lewontin, *ibid.*
- 41 Gould and Lewontin, 597.
- 42 De Waal, *ibid.*

HEGEL AND THE PATHOLOGIES OF FREEDOM¹

Paolo A. Bolaños
Macquarie University

“The history of the human world does not begin with the struggle between the individual and nature, since the individual is really a later product in human history. The community (*Allgemeinheit*) comes first, although in a ready-made, ‘immediate’ form. It is as yet not a rational community and does not have freedom as its quality. Consequently, it soon splits up into numerous antagonisms.” – Herbert Marcuse²

Recognizing the Self: Master-Slave Dialectic

One of the most important features of Hegel’s *Phenomenology of the Spirit* is his critique of the solipsistic tendency of Cartesian-Leibnizian ontological psychology.³ For Hegel, a “self-evident Self” is not readily available; the recognition of the Self requires more than Descartes’ *Cogito ergo Sum*. In Chapter IV of the *Phenomenology*, Hegel uses the notions of “desire” and “life-and-death” struggle to explain how such idea of a Self is even possible. What we find in the *Phenomenology* is an illustration of the ontological conditions from which “self-consciousness” is brought about. The movement from desire to the life-and-death struggle results in the Master-Slave dialectic, of which Hegel sees as part of further developments into what Alexandre Kojève calls “ideologies of Freedom.” It is, however, the contention of this paper that Hegel, himself, regarded these so called “ideologies” as deficient and, thus, “pathological.”

Marcuse writes that the “history of the human world does not begin with the struggle between the individual and nature, since the individual is only a later product in human history.” I would suppose that this is a reiteration of Hegel’s supposition that there is no self-evident Self. This means that the attainment of self-consciousness, for Hegel, is more like a journey, or better yet a *dialectical* progression, rather than a ready-made or transcendent idea that can be accessed through human reason. Further, when we say that the Self is not readily available, then we also have to accept that human rationality is also not self-evident. To my mind, the gist of Chapter IV of the *Phenomenology* is Hegel’s discussion of how self-consciousness originates from primitive states and how this self-consciousness becomes one with Absolute consciousness—that is, how human consciousness finally becomes Rational. Inasmuch as self-consciousness is not readily available, Hegel argues that it becomes available only through a series of struggles—that is, only through conflict. “Self-consciousness,” Hegel writes, “exists in and for itself when, and by the fact that, it so exists for another; that is, it exists only in being acknowledged.”⁴ In other words, the Self is not self-evident because it requires the existence of other selves to recognize it. “*Self-consciousness achieves its satisfaction only in another self-consciousness.*”⁵ The unrecognized individual is not, therefore, a Self. In a slight modification of Fichte, Hegel tells us that an individual must first prove himself before being recognized and, thus, must enter the realm of conflict and struggle⁶ or, in other words, a self-evident “I” presupposes an elemental *desire* to be a Self which then presupposes the

recognition of others. Accordingly, “desire points to a ‘doubling’ of self-consciousnesses, who are themselves not yet fully fledged mutually recognizing ‘self-consciousnesses’ but only egoistic proto-subjects . . .”⁷ The Master-Slave dialectic enters the picture precisely at this point of the *Phenomenology* wherein conflict is discussed.

Jean Hyppolite summarizes the movement toward self-consciousness clearly:

SELF-CONSCIOUSNESS, which is desire, can reach its truth only by finding another living self-consciousness. Three moments—the two self-consciousnesses posed in the element of externality and also externality itself, the Dasein of life—give rise to a dialectic which leads from the battle for recognition to the opposition between master and slave, and from that opposition to liberty.⁸

We notice a dialectical movement in the development of human self-consciousness: desire—master and slave—freedom of self-consciousness. My aim is to elucidate on the third moment, which is, for Kojève, the development of ideologies of freedom.⁹ I would personally refer to these as “pathologies of freedom.” As mentioned earlier, Hegel regards each specific moment (skepticism, Stoicism, and the unhappy consciousness) as deficient and, as such, are pathological. They are pathological inasmuch as they inspire excitement in the individual; each lends an individual or a group of people to embrace a deficient worldview that promises freedom. It is also important to note that, although Hegel talks of three major moments, each moment consists of series of moments. Perhaps, what Hegel wants to point out is that the internal or *immanent* conflicts found in each moment are the very conditions why the moments emerge. For instance, in desire, he talks of the dialectic between two self-consciousnesses a movement from one’s frustration with the self’s imposition of itself on objects. Objects do not recognize so we turn to other individuals for recognition. This interaction between two individuals results in the Master-Slave dialectic, which is explained by our attempt to impose ourselves on others. The imposition of selves or “proto-subjects”¹⁰ results from the elemental desire of an individual to assert himself which inevitably clashes with the desire of another, external, proto-subject to assert himself. A conflict arises because now we are dealing with live, and thus active/reactive, individuals and not merely passive objects. As “I try to impose my will on you, so you will try to impose your will on me: we will then end up in conflict,”¹¹ which leads to the life and death struggle. This conflict between master and slave leads to the contemplation of freedom—the slave’s emancipation from bondage to his master. For Hegel, however, this emancipation manifests when the slave finally succumbs to the master. Kojève writes: “before *realizing* Freedom, the Slave imagines a series of ideologies, by which he seeks to justify himself, to justify his slavery, to reconcile the *ideal* of Freedom with the fact of Slavery.”¹² From here, Hegel will talk about the three moments within the moment of the freedom of self-consciousness. We will notice that the Master-Slave dialectic is used as a kind of phenomenological narrative about the three moments (Stoicism, Skepticism, and Unhappy Consciousness).

Stoicism: the Slave and his Rational Idea of Freedom

The slave’s first reaction to the master is that of Stoicism. This form of ideology is manifested in the slave’s attempt to rationalize the apparently irrational world. Rationalizing, in this context, is the Stoic’s attempt to attune himself to the grand scheme of things. The phrase, “Man is a rational animal,” is taken seriously, even literally, here. Man is rational because he participates in a world governed by the *logos*. The Greek emphasis on thought/Reason has always been the medium by which we gain access to the *logos*. The Stoic ideal,

for Hegel, is the supposition “that consciousness is a being that *thinks*, and that consciousness holds something to be essentially important, or true and good only in so far as it *thinks* it to be such.”¹³ The Stoic’s ideal is that of honest work and virtue, and he convinces himself that he is free simply by having the abstract idea of freedom. This is a fleeing away from the actual conditions of the world; the Stoic deludes himself that he is autonomous, that he is not affected by the natural conditions of the world.

The manifold self-differentiating expanse of life, with all its detail and complexity, is the object on which desire and work operate. This manifold activity has now contracted into the simple positing of differences in the pure movement of thinking. Essential importance no longer attaches to the difference as a specific *thing*, or as consciousness of a specific *natural existence*, as a feeling, or as desire and its object, whether this is posited by myself or by alien consciousness. What alone has importance is the difference posited by *thought*, or the difference which from the very first is not distinct from myself.¹⁴

The movement from active bondage to Stoicism is, for Hegel, an ironic moment for the slave. The slave succumbs to the illusion of Reason in order to be liberated from bondage. The bondage is justified by the slave’s acceptance of his lot; accepting it, as it were part of the rational scheme of things. The slave rationalizes his inaction—he abandons the life and death struggle and merely submits to the master. It is tempting to interpret the master here as Reason itself, and as such we are reminded by Descartes and Kant who both submit to the rule of Reason. Moreover, I would see this as Hegel’s critique of Rationalism in general:

Freedom in thought has only *pure thought* as its truth, a truth lacking the fullness of life. Hence freedom in thought, too, is only the Notion of freedom, not the living reality of freedom itself.¹⁵

The realm of thought is prioritized and the real physical world is neglected. Hegel continues:

. . . here the Notion as an abstraction cuts itself off from the multiplicity of things, and thus has no content *in its own self* but one that is *given* to it. Consciousness does indeed destroy the content as an alien *immediacy* [*Sein*] when it *thinks* it; but the Notion is a *determinate* Notion, and this determinateness of the Notion is the alien element which is has within it.¹⁶

I would take the above as a moment when the Stoic imposes an artificial truth (a truthful lie for Nietzsche) upon a multiplicity of things we call the world. Like a schizoid, the Stoic constructs for himself an imaginary world. Hegel’s critique of the “True and the Good”¹⁷ might as well be Plato’s; but, perhaps, this might be an allusion to Kant who argues that metaphysical ideas are ideas of reason and are totally remote from sensibility.¹⁸ “But this self-identity of thought is again only the pure form in which nothing is determined.”¹⁹ Because nothing is determined in Stoic attitude, or at least for Hegel it is an incomplete movement of the slave’s consciousness, it should pave the way to a new ideology Hegel calls skepticism.

Skepticism: the Nihilistic Slave

The passivity of the Stoic and the self-identity of reason lead to boredom.²⁰ Eventually, the Stoic gets disillusioned by his discovery that virtue is not always rewarded and honest

work is often exploited. The next stage, Skepticism, is the wanting to be active again—the “realization of which Stoicism was only the Notion, and is the actual experience of what the freedom of thought is.”²¹ This is a movement back to the friction between master and slave. The slave has awakened from its formalistic slumber and is now ready to annihilate previous suppositions and to start anew. By becoming active the slave can again negate the real and, thus, change it. The Skeptic disdains the ineptness of the slave to fight the master and he succumbs to sheer cynicism of the meek way of life. This, however, results to the denial of everything real.

In Skepticism, now, the wholly unessential and non-independent character of this ‘other’ becomes explicit *for consciousness*; the [abstract] thought becomes the concrete thinking which annihilates the being of the world in all its manifold determinateness . . .²²

If the Stoic accepts the Master-Slave relationship by rationalizing it, the Skeptic, for his part, takes action and denies this relationship; this is done through the complete denial of otherness. The total negation of everything real, however, ends up in Solipsism²³—which is a reversion to the Cartesian *Cogito*. At the beginning of Chapter IV of the *Phenomenology* Hegel criticizes this tendency towards solipsism and argues that self-consciousness is only possible in the eyes of another self-conscious being. Hence, a total negation of reality would also imply the negation of other self-consciousnesses. This implies that to deny others is to deny oneself. Kojève comments that “to take nihilism seriously is to commit suicide, to cease completely to act and—consequently—to live.”²⁴ Skepticism is schizophrenia without content—there is no possibility of the world for the Skeptic. Hegel writes:

The skeptical self-consciousness thus experiences in the flux of all that would stand secure before it its own freedom as given and preserved by itself . . . It affirms the nullity of seeing, hearing, etc., yet it is itself seeing, hearing, etc. It affirms the nullity of ethical principles, and lets its conduct be governed by these very principles.²⁵

For Hegel, Skepticism is seen as a withdrawal from reality—it is an escapist attitude. The Skeptical view contradicts itself in its acceptance of the empirical world, yet denying the intelligibility of such world; the Skeptic depends on what he negates. It is from this nihilistic attitude towards the world that consciousness develops into another level.

Unhappy Consciousness: the Religious Slave

Hegel thinks that the slave’s desire for freedom persists. Thus, he talks of another development in self-consciousness, that of the Unhappy Consciousness. From the rationalizing Stoic to the nihilist Skeptic, the slave enters the world of Unhappy Consciousness with the master; this is depicted in Christian ideology. The Stoic attitude is delusory, the Skeptic is evasive; meanwhile, Unhappy Consciousness

. . . is in fact *one* consciousness which contains within itself these two modes. This new form is, therefore, one which *knows* that it is the dual consciousness of itself, as self-liberating, unchangeable, and self-identical, and as bewildering and self-perverting, and it is the awareness of this self-contradictory nature itself.²⁶

The Unhappy Consciousness is a duplication of the master and the slave in one—“the duplication which formerly was divided between two individuals, the lord and the bondsman, is now lodged in one.”²⁷ Kojève tells us that the

Slave does not deny the contradictory character of his existence. But he tries to justify it by saying that *all* existence necessarily, inevitably implies a contradiction. To this end he imagines an “outer world,” which is beyond (*Jenseits*) the natural World of the senses.²⁸

The unhappy conscious individual takes the Skeptic’s nihilism a step further: he completely withdraws from the world and becomes self-absorbed in prayer. This development has a strong resemblance to what Nietzsche calls the ascetic ideal. It is a movement away from the Stoic and the Skeptic attitudes of masking and avoiding the being of the world. The irrational and chaotic world of change is rejected; but, then again, with the slave’s act of rejecting the world of flux, he posits a beyond, he creates a religion. The slave posits a beyond because it is still unable to reach the concrete identity of certainty and truth.²⁹ As a side effect to the pronouncement of a beyond, the world is bifurcated (ala Plato). A dualistic world comprising the contingent and the eternal arises: “the simple Unchangeable, it takes to be *essential* Being; but the other, the protean Changeable, it takes to be the unessential.”³⁰ The slave accepts his life of strife (his contingent life) in the hope that a life after death is available (promise of the eternal). “This unhappy consciousness is *subjectivity*, which aspires to the repose of unity; it is self-consciousness as consciousness of life and of what exceeds life.”³¹ This is a notion of freedom removed from Stoicism and Skepticism; freedom here is transcendent, freedom is in the beyond. The goal, therefore, of the slave is to free himself from the unessential (mutable/human finitude) and ascend to the essential (immutable). This ideology of freedom imposes on man a dual nature: man is at the same time mutable and immutable. It is an ideology based on the belief that we are all created in the image and likeness of a divine God and, thus, we are also divine beings.

The ascent to the divine God happens in several ways: prayer (pure consciousness), work (desire and work), and penitence (consciousness that is aware of its own being-for-itself).³² These three are, for Hegel, the basic ideals of Christian ideology. Prayer is a kind of search for the divine God; the religious person finds communion with the beyond. Work is the ideal of trying to serve God through labor. The world, in this context, is the space for labor; the slave harnesses nature to become closer to God.³³ One also has to renounce oneself through penitence, the gesture of regret expressed through torture or even mutilation of the body. These forms for punishment suppose to show that the body is insignificant while the spirit is of the utmost importance. The act of torturing the body (through whipping or crucifixion) is a symbolic gesture of imitating the passion of Jesus; this symbolizes sympathy towards the Christ by taking on his burden of carrying the cross and being nailed unto it.

Hegel criticizes prayer for its ultra-emotive character, work for it poses a contradiction between the finite world and God, and penitence for it demotes the body into one of weakness and corruption.³⁴ Hegel also observes that the Unhappy Consciousness gives up all its desires and freedom by submitting to a mediator (the priest). “This mediator, having a direct relationship with the unchangeable Being, ministers by giving advice on what is right.”³⁵ To my mind, however, this submission to the priest is precisely the very essence of self annihilation; and thus the goal of self-consciousness turns against itself.

Conclusion

Although I refer to the three ideologies as pathologies of freedom, the Hegelian point is that in order to climb up the ladder, human consciousness did experience these pathologies. In other words, a series of *Aufhebungen* did occur. The Unhappy Consciousness is not yet the culmination of the Spirit, for it fails to mediate self-consciousness with the universal. As

mentioned, self-consciousness turns against and negates itself by its total submission to the universal via the priest; thus, real freedom is not achieved. The Unhappy Consciousness, albeit pathological, is a *conditio sine qua non* for the next level: the dialectic of Reason.

Meanwhile, Deleuze thinks that Hegel's supposition of the Absolute spirit has only fallen back to the very metaphysics of transcendence that the latter sought to abandon.³⁶ We have seen that even if Hegel criticizes the Unhappy Consciousness, he seems to accept this moment inasmuch as one "thinks" of and with the Absolute, and thus becomes closer to (Absolute) Spirit. I wonder whether Hegel is talking about an Absolute Spirit or simply the spirit of God—is there a difference? There are, of course, at least two schools of thought regarding this very aspect of Hegel's philosophy. On the one hand, right-wing Hegelians are inclined toward a conservative, and rather theological, interpretation of the notion of the Absolute; they sought to reconcile it with Christianity. On the other hand, the left-wing Hegelians are inclined toward a more political and atheistic account of the Absolute. It seems that today Hegel's audience consists of critical theorists who are more sensitive to the aptness of Hegelian dialectics as a basis for a critique of society.

Works Cited:

- Deleuze, Gilles, *Difference and Repetition*, trans. by Paul Patton (New York: Columbia University Press, 1994).
- Hegel, G.W.F., *Phenomenology of the Spirit*, trans. by A.V. Miller (New York: Oxford University Press, 1977).
- Hypolite, Jean, *Genesis and Structure of Hegel's Phenomenology of Spirit*, trans. by Samuel Cherniak and John Heckman (Evanston: Northwestern University Press, 1974).
- Kant, Immanuel, *Critique of Pure Reason*, trans. by Norman Kemp Smith (London: MacMillan and Company, Ltd, 1960).
- Kojeve, Alexandre, *Introduction to the Reading of Hegel: Lectures on the Phenomenology of Spirit*, trans. by James H. Nichols, Jr. (Ithaca: Cornell University Press, 1969).
- Marcuse, Herbert, *Reason and Revolution: Hegel and the Rise of Social Theory* (New York: Oxford University Press, 1960).
- Sinnerbrink, Robert, "Recognitive Freedom: Hegel and the Problem of Recognition," in *Critical Horizons*, 5 (2004), 272-295.
- Solomon, Robert, *In the Spirit of Hegel: A Study of G.W.F. Hegel's Phenomenology of Spirit* (New York: Oxford University Press, 1983).
- Stern, Robert, *Hegel and the Phenomenology of Spirit* (London: Routledge, 2002).
- Weber, Max, *The Protestant Ethic and the Spirit of Capitalism*, trans. by Talcott Parsons (London: Routledge, 1992).

Footnotes:

- 1 I would like to thank Robert Sinnerbrink for his invaluable comments on an earlier draft of this paper.
- 2 *Reason and Revolution: Hegel and the Rise of Social Theory* (New York: Oxford University Press, 1960), 73-74.
- 3 Robert Solomon discusses this in detail in *In the Spirit of Hegel: A Study of G.W.F. Hegel's Phenomenology of Spirit* (New York: Oxford University Press, 1983), 428-ff. 4 G.W.F. Hegel, *Phenomenology of the Spirit*, trans. by A.V. Miller (New York: Oxford University Press, 1977), §178. Robert Sinnerbrink's essay is particularly helpful with regard to the "recognitive" aspect of the dialectics of self-consciousness. See "Recognitive Freedom: Hegel and the Problem of Recognition," in *Critical Horizons*, 5 (2004), 272-295.

- 5 Hegel, *op cit.*, §175.
- 6 Cf. Solomon, *op cit.*, 433-434.
- 7 Sinnerbrink, *op cit.*, 276.
- 8 Jean Hyppolite, *Genesis and Structure of Hegel's Phenomenology of Spirit*, trans. by Samuel Cherniak and John Heckman (Evanston: Northwestern University Press, 1974), 156.
- 9 See Alexandre Kojève, *Introduction to the Reading of Hegel: Lectures on the Phenomenology of Spirit*, trans. by James H. Nichols, Jr. (Ithaca: Cornell University Press, 1969), 53.
- 10 Cf. Sinnerbrink, *op cit.*, 276-ff.
- 11 Robert Stern, *Hegel and the Phenomenology of Spirit* (London: Routledge, 2002), 76.
- 12 Kojève, *op cit.*, 53.
- 13 Hegel, *op cit.*, §198.
- 14 *Ibid.*, §199.
- 15 *Ibid.*, §200.
- 16 *Ibid.*
- 17 *Ibid.*
- 18 See for example Kant's supposition that the "pure concept, in so far as it has its origin in the understanding alone (not in the pure image of sensibility), is called a *notion*. A concept formed from notions and transcending the possibility of experience is an *idea* or concept of reason. *Critique of Pure Reason*, trans. by Norman Kemp Smith (London: MacMillan and Company, Ltd, 1960), 314.
- 19 Hegel, *op cit.*, §200.
- 20 See Kojève, *op cit.*, 53-54.
- 21 Hegel, *op cit.*, §202.
- 22 *Ibid.*
- 23 Cf. Kojève, *op cit.*, 54.
- 24 *Ibid.*
- 25 Hegel, *op cit.*, §205.
- 26 *Ibid.*, §206.
- 27 *Ibid.*
- 28 Kojève, *op cit.*, 55.
- 29 Cf. Hyppolite, *op cit.*, 190.
- 30 Hegel, *op cit.*, §208.
- 31 Hyppolite, *op cit.*, 195.
- 32 See Hegel, *op cit.*, §214.
- 33 Max Weber makes a similar point and poses a challenge to Marxism. He argues that the development of capitalism was strongly influenced by religious ideals, in particular Christian religious ideals. See *The Protestant Ethic and the Spirit of Capitalism*, trans. by Talcott Parsons (London: Routledge, 1992).
- 34 For a more detailed discussion see Stern, *op cit.*, 93-94.
- 35 Hegel, *op cit.*, §228.
- 36 Deleuze criticizes Hegel for declaring that the negating dialectic is the motor of history (as shown above), which signals the idea that through a series of *Aufhebungen* history will finally end one way or another—that is the realization of the Absolute Spirit. Deleuze, however, loathes the idea that history has a teleological nature. "History progresses not by negation and the negation of negation, but by deciding problems and affirming differences." *Difference and Repetition*, trans. by Paul Patton (New York: Columbia University Press, 1994), 268.

Where Mary went wrong

Cressida Gaukroger

The knowledge argument against physicalism rests on the belief that we can conceive of a person who knows all the physical facts about the world, and yet still does not know all the facts (most importantly, all the psychological facts about the world). Physicalism is the thesis that psychological truths supervene on physical truths, so to show that knowledge of psychological truths does not necessarily follow from knowledge of physical truths is to undermine the strength of their connection central to the physicalist's doctrine. I argue in this paper that this argument against physicalism is brought into question when we look at the most famous knowledge argument – Frank Jackson's creation of Mary in the black and white room.¹ In Jackson's story, the protagonist, Mary, appears to know all the physical facts about the world and yet she has not had phenomenal experiences of colour sensations. I argue that there are some physical facts which rely on phenomenal experiences to be knowable, and that what this version of the knowledge argument really shows is that *Mary cannot know all the physical facts*, which in fact can offer support to physicalism as opposed to proving it wrong.

Frank Jackson's Knowledge Argument

Mary has spent her life confined to a black and white room. Everything she sees in this room is seen as if looking at a black and white television. She has never experienced a colour sensation. Mary also happens to be a brilliant scientist who has learnt *all* the physical facts about the world, and yet when she suddenly receives a bright red apple in the post, she appears to learn something new – something about psychological experiences (both her own and those of others).

This can be seen as an argument against physicalism in two different ways.² First, it can be concluded that if Mary knows all the physical facts, and yet there is still something that she doesn't know, then this means that there is more to the world than what can be captured in terms of physical facts.

Secondly, it attacks the psychophysical conditional at the heart of physicalism which can be expressed in the following terms. If P represents all the physical truths of the world and P* represents all the psychological (phenomenal) truths of the world then it is necessarily true that:

If P then P*.³

The knowledge argument shows a logically conceivable situation where someone knows all the physical truths and yet still lacks knowledge of the psychological (phenomenal) truths, meaning that the antecedent of the psychophysical conditional can be true and the consequent false. The best explanation of this is that physicalism is false or, at most, contingent.⁴

The knowledge argument in this form assumes the following premises:

1. In the black and white room, Mary knows all the physical facts there are to know.
2. In the black and white room, Mary does not appear to know all the psychological (phenomenal) facts there are to know. (This is shown by the fact that she seems to learn some of these second kind of facts when she experiences colour for the first time).

Therefore,

3. The black and white room example shows us that the psychological (phenomenal) facts do not supervene on the physical facts, and (probably because) they are *different* to physical facts – there is *more* to know than just the physical facts about the world.

While most responses to the knowledge argument focus on the second premise and the conclusion, I argue in this paper that it is the first premise that is at fault.⁵

§1. What *Can* Mary Know?

In the black and white room, Mary knows *all* the physical facts about the world. Or does she? When she first receives a bright red apple – when she first sees it and has her first red experience, she learns something new. And this ‘something new’ does not just seem to be knowledge about her own experiences “so *this* is what it is like to see red” or the experiences of others “so *this* is what they see when they see red/red apples”, but I would argue that she also learns something new about *apples*.

Mary has seen what an apple looks like (in black and white), she has learnt that wavelength combinations reflected from the surfaces of apples cause one of a couple of possible brain states in the people with normal eyesight who observe them, which lead these people to say that they are seeing red, or that the apple in question is red. And yet there is something about the apple that she does not seem to know until she sees it in full colour.

When other people have told Mary that apples are red⁶ (or that the experience that one has when they look at an apple is an experience of seeing *redness*), the way that she understood what they meant was different *before* she experienced red herself to the way she understood it *after* having such an experience. In fact, before she saw the colour red, the proposition “apples are red” would have been a completely different proposition from the proposition “apples are red” as she knows it now that she has experienced red. Put in another way, what a colour-blind person means when they utter the sentence “apples are red” is not the same as what a normal-sighted person means when they utter the sentence “apples are red” because the two propositions they express involve different concepts of what it is for something to be red.

In the black and white room, Mary is missing certain *concepts* that are necessary components in understanding situations where those concepts are applied, such as when a normal-sighted person says “apples are red”. The missing concepts are *phenomenal concepts* and in this case the concept that Mary is at first missing, but learns upon seeing the red apple, is the phenomenal concept *red sensation*. It is these types of concepts, phenomenal concepts, that appear to be what is referred to when we conclude from this example that there is something more than just physical concepts – and we come to this conclusion because knowledge of phenomenal concepts cannot be derived from knowledge of physical facts, we need to experience a phenomenal concept to be able to come to hold it. Mary does appear to learn new phenomenal concepts when she leaves the black and white room, but *this is not all she learns*.

By having the phenomenal concept *red sensation*, Mary can now apply her new knowledge and discover other new things such as “this is what it is like to have a red sensation”, “this is what others experience when they look at apples”, “red is brighter than grey” and “this is what everyone else meant when they said that ‘apples are red’.” She now knows “apples are red” in the way that normal sighted people know it, and she realises that when she had previously claimed to know the fact “apples are red” she had meant something quite different by it. This could well be what Jackson means when he says that Mary learns “something about the world”.⁷

But, what Mary learns about the world, that “apples are red” and a host of other propositions opened up to her by her newly acquired possession of certain phenomenal concepts (phenomenal concepts regarding colour sensations to be precise), *are physical facts*. That “apples are red” is a *physical fact*. Similarly, there are many physical facts that require possession of phenomenal concepts for us to know them – “poison ivy is painful to touch,” “pineapple tastes sweet”, “off milk smells rancid”, “the musical note ‘A’ sounds different to the musical note ‘D’” – so it should be not be such a surprise that, missing certain relevant phenomenal concepts, Mary, whilst in the black and white room, is not able to know all the physical facts.

Hence, the founding premise upon which Jackson’s knowledge argument is based is false. Jackson’s argument does *not* stand as an example of someone who knows all the physical facts, but does not know all the facts. One might argue that it does not follow that it would be *impossible* for someone to know all the physical facts but not all the phenomenal facts - just because there are some physical facts that are unknowable without the relevant phenomenal concepts does not entail that no phenomenal concepts are entirely independent of physical facts so that knowledge of them (or absence thereof) would not have any bearing on our knowledge of any physical facts. However, I would challenge anyone to devise such an example which identified such independent phenomenal concepts. This example reveals that even the seemingly independent cases of phenomenal concepts (as in the redness of apples) are still intrinsically linked to physical facts, knowledge of which relies on our experience of such phenomenal concepts.

§2. Why Mary Knew Less Than We Thought.

I take knowledge here to be (at least) justified true belief. For Mary to know that apples are red, she must not only be *justified* in believing that apples are red – she must actually hold the specific belief that “apples are red”. Now, in a sense, Mary knows that apples are red because she knows the physical account of what happens to us when we see an apple, from the wave-lengths that are given off light hitting the apple to our saying “apples are red”. But this is not the sense in which ordinary competent speakers of English mean that apples are red. To know that apples are red in the common way that it is known, is to know a fact that relies on a knowledge of the phenomenal concept involved, in this case the phenomenal concept *red sensation*. Because, while in the black and white room, Mary cannot conceive of what it means to normal sighted people for something to have the quality of being red, she can hardly have the *belief* that “apples are red” (as it is known to normal sighted people) let alone the *knowledge* of this physical fact.

2.2 Marianna’s mistake and how it shows that Mary doesn’t understand.

Martine Nida-Rümelin presents a modified version of the Mary argument that is helpful in illustrating the way in which she cannot hold, what Nida-Rümelin calls, “phenomenal beliefs” about colour facts.⁸ Marianna, like Mary, had spent her life in a black and white room, and let’s say that, like Mary, she also knows all the physical facts about colour – perhaps all the physical facts about the world. One day Marianna is presented with four slides “showing clear cases of blue, red, green and yellow, and she is asked which of the four slides shows a clear case of the colour she believes to be experienced by normal people when they look at the sky.”⁹ She is not told the colour of any of the slides. Now Marianna, knowing all the colour facts, knows that the colour of the sky is often heralded as bright, dazzling and brilliant, and when she sees the red slide it captures her eye as having all these qualities. So, she immediately forms the belief that the colour normal-visioned people (such

as herself) see when they look at the sky is the colour that she sees when she looks at the red slide.

Marianna knows what it is like to see red because she is acquainted with the colour of red. But she does not know anything else about this colour, because she does not know that it is what we refer to when we refer to the colour red. When she is first told that the sky is blue, we may assume (as we originally did in the Mary case) that she has learnt that the sky is blue. But the mistake that she makes when trying to pick out the colour blue from other colours shows us that she actually lacks the concept of what it is for something to be blue. When shown the colours blue and red she forms the belief that the sky is red *because she thinks that blue appears *red* to normal people*. Even when she has been acquainted with phenomenal concepts, for Marianna what it means for something to be blue is different to what it means to us for a thing to be blue, which means that her belief that the sky is blue is different to our belief that the sky is blue. This is because she is mistaken – she does not actually know that the sky is blue.¹⁰

One may argue that the Marianna case is merely based on a naming mistake and so does not tell us any more about Mary's situation. But the fact that Marianna can be mistaken in such a way about the colour of the sky shows something important about her initial belief about the colour of the sky. If either woman has no knowledge of what it is like to have a blue sensation (or no knowledge of what *is* a blue sensation) then there is a sense in which they do not *know* that the sky is blue. The possibility of a simple naming error just shows that Marianna, at least before she has any colour experiences, knows that the sky is blue only in the sense that blue is an empty term which designates nothing which she can properly understand.

Another response is that Marianna does not even have all the physical facts about the world that she can understand – for if she had *all* the physical facts, she would also know the names of the coloured slides that she was seeing, because “this slide is red” is a physical fact (and one that would then qualify her phenomenal concept *red experience*). It is unclear, however, how this takes away from the example. In the original example it is not necessary for Marianna to have a complete knowledge of physiology because her case is used “only to introduce an epistemological distinction” and not as an argument against physicalism.¹¹ Similarly, I use it here to illustrate the point that Mary cannot possibly understand what we mean by “red” or “blue” – and that knowledge of phenomenal concepts is necessary (although, we have seen, not sufficient) for her to know a physical fact such as “the sky is blue”.

2.3 Scientific facts and why justification is not enough

Another response to my argument is that there are lots of facts (scientific facts in particular) that we can know even without direct access to them because we can gain knowledge of them by deferral to experts. We do not need to work out for ourselves that water = H₂O, for instance, nor do we need a complete grasp of chemistry to know it – we know that water = H₂O because we are told it by scientists, and that is justification enough to make it knowledge. What is more, we have learnt these kind of facts in exactly the same way that Mary has learnt that apples are red and the sky is blue. What makes the two cases different?

The answer to this is that these cases are actually not so different because even in the case of scientific facts, such as water = H₂O, one needs to have a certain grasp of the concepts involved (water, H₂O, and what it means for them to be one and the same thing) for it to count as knowledge. I will use the example of the fictional and ignorant Igor to illustrate this point.

Igor has been told that water = H₂O. He can repeat this fact, he can point at water and say “H₂O” and so on. Does this mean that he actually knows that water = H₂O? It turns out that Igor has *no idea* what it means for something to be H₂O. Even if he knew that there were such things as molecules, which we can assume that he doesn’t, he would still have no clue that the term “H₂O” refers to molecules etc. In this way, what Igor knows is that there is some other term that we use to refer to water, because *all* he knows about this term is that it refers to water. But that “water has many names” is not the substance of the proposition “water = H₂O” – the kind of understanding that Igor has does not capture this fact at all.

When we say that x knows that water = H₂O, we assume that x has some knowledge of what H₂O is other than it being one and the same as water – otherwise the *meaning* of the sentence “water = H₂O” to them is nothing more than “water = water” (and apart from anything else, this should be knowable *a priori*, and so we would not need deferral to experts). If someone has no conception of what it is to be H₂O, then they can hardly *understand* the fact that it is water and understanding does appear to be necessary in this case for knowledge to be present.

The obvious response to Igor’s case is to say that his position is in no way the equivalent of Mary’s. She not only knows that the “sky is blue” in the same way that Igor knows that “water = H₂O” (that is to say, she can repeat this fact to others, and she can point a picture of the sky and say “blue”), but, while she may not know everything there is to know about the colour blue (namely, what it looks like) she knows *a lot* about blue – which makes her case very different to Igor’s, where he knows *nothing* about H₂O. In this instance, the differences between the cases become apparent when we consider the kind of things that Igor needs to know before we can say that he knows what it means for something to be H₂O, and the kinds of things that Mary needs to know before we can say that she knows what it means for something to be blue. Igor needs to know some physical facts about what molecules are and how H₂O relates to elements and their role in the makeup of substances before he can understand the fact water = H₂O.

The kinds of things Mary needs to know, however, for her to understand what it means for something to be red are not only physical facts, as they are in Igor’s case – Mary’s knowledge of the physical facts surrounding the meaning of the term “red” are not sufficient for her to have knowledge of what it means for something to be red. In fact, if we do not accept this point, then we would have to also reject the conclusion of the knowledge argument, because it specifically rests on the premise that the possession of knowledge of physical facts is not sufficient for facts that include phenomenal concepts, such as what it means for something to be red. If the kind of knowledge that we have about what it means to be H₂O was exactly the same kind of knowledge needed for us to understand what it means to be red, then Mary shouldn’t experience such a revelation when she sees red for the first time.

§3. Does Mary Know Enough Even Without Knowing Everything?

Another possible response to my argument is the question “does Mary need to know *all* the physical truths for the knowledge argument to be effective?” In the original, 1982 version of the Mary knowledge argument, Jackson’s Mary “specializes in the neurophysiology of vision and acquires, let us suppose, all the physical information there is to obtain about what goes on when we see ripe tomatoes, or the sky, and use terms like ‘red’, ‘blue’, and so on.”¹² She does not know all the physical facts that one could possibly know about the world.

In Jackson’s slightly later account of Mary, she appears to know more: “...she learns

everything there is to know about the physical nature of the world. She knows all the physical facts about us and our environment, in a wide sense of ‘physical’ which includes everything in *completed* physics, chemistry, and neurophysiology, and all there is to know about the causal and relational facts consequent upon all this, including of course functional roles.”¹³ Facts such as “apples are red”, “poison ivy is painful to touch”, “pineapple tastes sweet”, “off milk smells rancid” and “the musical note ‘A’ sounds different to the musical note ‘D’”, may be included in the things that Mary is supposed to know when she is in the black and white room of Jackson’s 1986 description, but nothing in the 1982 description suggests that she must know these facts. Let’s say that Mary does not know all the physical facts – she is missing those facts that directly rely on phenomenal concepts that she does not possess, such as the concept of *red sensation* – how much does this actually affect the conclusion of the knowledge argument?

If Mary does not know all the physical facts that can possibly be known, then we do have to give up what I have called the second conclusion of the knowledge argument. This conclusion of the knowledge argument is that, because it is conceivable that someone could know all the physical facts, but not know the psychological (phenomenal) facts, the antecedent of the psychophysical conditional can logically be true, while its consequent is false, and therefore the psychophysical conditional can be best explained as either false or contingent. What I have shown is, however, that because possession of relevant phenomenal concepts is necessary for knowledge of some physical facts, we *cannot* conceive of a situation where someone knows all the physical facts, but does not know all the psychological (phenomenal) facts.

The first conclusion of the knowledge argument, however – the thesis that if Mary knows all the physical facts, and yet there is still something that she doesn’t know, then there must be more to the world than what can be captured in physical facts – can arguably still be salvaged by being replaced with a slightly weaker thesis that leads us to the same conclusion. The fact that Mary appears to know the full physical account of what happens when someone sees colour, and yet she lacks the full knowledge of what happens when one sees colour, would give us the conclusion that there is more to know about colour vision than can be captured in physical facts alone, and, therefore, more to know about the mind than can be captured in physical facts alone because colour vision is an important aspect of the mind.

A weaker thesis may be something such as: for some particular aspect of the mind, if Q is the full physical account of this aspect of the mind and Q* is the psychological (phenomenal) account of this aspect of the mind, then it is necessary that:

If Q then Q*.

Mary can know the full physical account of what happens when we see colour, and yet still lack the psychological (phenomenal) account of what happens. There is, therefore, more to know about colour vision than can be captured in a physical account of the mind, and consequently, more to know about the mind than can be captured in a full physical account of the mind.

In fact, this ‘something more’ that Mary misses in her understanding of colour vision is something I have not tried to deny when presenting my response to the knowledge argument. What I argue is not that if Mary knew *all* the physical facts she would therefore have a full understanding of phenomenal concepts.¹⁴ Phenomenal concepts do have qualities that make their possession different from the possession of other concepts. The most striking difference between phenomenal concepts and other concepts is that one

needs to have experience of phenomenal concepts to be able to possess them. This understanding of phenomenal concepts is called the experience thesis, and is held by philosophers such as David Papineau and Michael Tye.¹⁵

The experience thesis also explains why there are some things that Mary cannot know in the black and white room (provided that she does not dream of red, or her brain is not operated upon so that she goes into the *red sensation* brain state etc. – for these are all ways of experiencing, and therefore, gaining, the phenomenal concept *red sensation*). While I argue that what Mary would *need* to have if she were to know all the physical facts is an experience of relevant phenomenal concepts, I am still bound to explaining the nature of these phenomenal concepts and why they are not just included in, or deducible from, physical facts.

All the weakened conclusion of the knowledge argument can achieve, therefore, is the acknowledgement that phenomenal concepts exist that can not be deducible from physical facts. The existence of phenomenal concepts does not, however, necessarily pose a problem for physicalism. What happens to Mary when she learns to possess phenomenal concepts of colour sensations can still be explained in physical terms – upon seeing red for the first time, she finally experienced the being in brain-state that, as she had always known, causes a *red sensation*. The fact that experience is necessary for possession of phenomenal concepts also explains why she could not deduce phenomenal concepts from physical facts. Hence physicalism can explain away the apparent contradiction of phenomenal concepts.

Mary must, therefore, know *all* the physical facts about the world while she is in the black and white room for the knowledge argument to mount a proper challenge to physicalism. My response to the knowledge argument, that Mary *cannot* know all the physical facts, remains, therefore, a strong one.

§4. *Physical Facts or Phenomenal Facts?*

Martine Nida-Rümelin uses her example of Marianna (see §2.2) in a different way to the way I have used her. For Nida-Rümelin, Marianna’s mistake is used to highlight an epistemological distinction between phenomenal and non-phenomenal belief.¹⁶ When Marianna is told that the sky appears blue to normal-sighted people, she believes it non-phenomenally, whereas when she sees the red slide which she mistakes as blue, her belief that what people see when they see the sky is what she sees when she sees the red slide, so her belief that the sky appears red to normal sighted people, is a phenomenal belief. This distinction between phenomenal and non-phenomenal beliefs is what explains Marianna’s otherwise apparently contrary beliefs. It could be concluded, therefore, that just as there are phenomenal and non-phenomenal beliefs, there can be phenomenal and non-phenomenal facts, and physical facts are only those captured in the latter category.

The final possible problem with my response to the knowledge argument that I will discuss here is the question of whether what I am concluding to be physical facts, “apples are red”, “the sky is blue” etc. are not actually physical facts, but something different, phenomenal facts, for instance.

Daniel Stoljar, in a defence of the knowledge argument from the ‘phenomenal concept strategy,’¹⁷ poses a modified Mary example which he calls “Experienced Mary” that runs along similar lines to Nida-Rümelin’s Marianna example.¹⁸ Experienced Mary, like Jackson’s Mary, has grown up in a black and white room and has a complete physical knowledge. Experienced Mary manages to escape and have colour experiences which allows her to possess the phenomenal concepts that she was missing in the black and white room. Unlike Jackson’s Mary, however, Experienced Mary not only gets recaptured and returned to the

black and white room, but, once back inside the room, she forgets something. She remembers what red sensations are like, for instance, so she doesn't lose any phenomenal concepts. And, Stoljar argues, "she loses none of the purely physical knowledge that she knew so impressively before her release."¹⁹ What Experienced Mary does forget is "in which cases phenomenal concepts [of colour sensations] are correctly applied. Thus what she forgets are truths of the form 'Fred's arthritis causes him pain', and 'Looking at Granny Smith apples causes green sensations'."²⁰

The question that this example raises is whether what Mary has forgotten is some of her physical knowledge or some of her 'phenomenal knowledge'. Stoljar assumes that Mary has forgotten some of her phenomenal knowledge (although not her knowledge of phenomenal concepts – "phenomenal knowledge of another sort").²¹ If this is correct, then my response to the knowledge argument has been based on the flawed assumption that there are some physical facts that one needs relevant phenomenal concepts in order to understand, when actually these facts such as "apples are red", "poison ivy is painful to touch," "pineapple tastes sweet" etc. are not physical facts after all, but *phenomenal* facts.

When considering the existence, and role, of phenomenal facts it is important that we do not just assume, as most people have done previously, that Mary *definitely* knows *all* the physical facts when she is in the black and white room, just *because* this is one of the main premises of Jackson's argument. If we assume that Mary knows *all* the physical facts there are to know while she is in the black and white room, then it does follow that anything she does not know in the black and white room *cannot* be a physical fact. However, this explanation is incomplete. To show that Mary cannot learn new physical facts when she leaves the black and white room we need an explanation of why facts such as "Fred's arthritis causes him pain", "looking at Granny Smith apples causes green sensations", "pineapple tastes sweet" and "off milk smells rancid" are not physical facts.

It seems like the most obvious response to this challenge is to point out that possession of phenomenal concepts is necessary for knowledge of the above facts, whereas it is not necessary for knowledge of physical facts. The above facts are, therefore, different from physical facts in the way that they rely on phenomenal concepts, and they are therefore *phenomenal* facts.

This conclusion becomes problematic, however, when we consider that possession of *some* phenomenal concepts, at least, is necessary for any knowledge at all, let alone knowledge of physical facts. Can we conceive of a case where someone possesses *no* phenomenal concepts and yet still has knowledge of physical facts? No. Apart from the obvious objection of how they would learn anything, it seems impossible to explain how they would even be able to conceive of anything, for what concepts would be available for them to do so? It would not take enough to just imagine a thinking 'brain in a vat'. The 'brain in a vat' would not have experienced *any* of the brain-states that consist of, or lead, in any way, to phenomenal sensations. When they 'think' they would not do so in aural, visual or any other terms that we think in. And even with the most basic abstract concepts such as the numbers 1 or 2, it is very hard to know how they would be able to develop or possess such concepts without a concept of 'thing' as understanding of these numbers seems to rely on an understanding of one 'thing' or two 'things'. Even knowledge of the concept 'nothingness' can only be conceived in contrast to the concept of 'something'.

The fact that *some* phenomenal concepts are necessary for knowledge of physical facts is not particularly new or contentious. It is a very popular view that phenomenal concepts are necessary for consciousness,²² and the view that consciousness is necessary for understanding, and therefore knowledge, is also pretty uncontroversial.²³ However, this means that facts that rely on the possession of phenomenal concepts to be known cannot

be different from physical facts as phenomenal concept possession is necessary for the knowledge of *any* facts whatever.

In response it could be pointed out that the facts that Mary does not know in the black and white room not only rely far more heavily on possession of phenomenal concepts than physical facts, but they have actually have a phenomenal *component* and it is this which explains the difference between phenomenal facts and physical facts. If this is true, the fact that “apples are red” has the phenomenal concept *red sensation* as one of its components in a way that a physical fact such as “water = H₂O” does not, and this is what sets “apples are red” apart as a phenomenal fact. However, if it is the case that all facts that have a phenomenal component are not physical facts but phenomenal facts, and it is also the case that phenomenal facts are facts that can only be learnt through experience, meaning that all phenomenal facts are *a posteriori* knowable, then the psychophysical conditional which is a fact that has a phenomenal component (namely the consequent), must be a phenomenal fact, and so must also be knowable *a posteriori*.

So, even if we do end up concluding that I have been wrong in claiming that Mary does not know all the physical facts, and actually what she learns are phenomenal facts, the psycho-physical conditional is not undermined by this conclusion. We have shown that, if there are phenomenal facts which are defined as such by a phenomenal component (the experience thesis telling us that such facts are only knowable *a posteriori*), then we can show that the psychophysical conditional is *a posteriori* and yet can not only retain its truth and necessity, but any appearance of contingency can also be explained away by the experience thesis.²⁴

Conclusion:

I began this paper by arguing that Frank Jackson’s black and white room version of the knowledge argument against physicalism is flawed as it tries to show that there could be someone who knew *all* the physical facts of the world, and yet still lacked knowledge of phenomenal concepts, when actually knowledge of relevant phenomenal concepts is necessary for knowledge of some physical facts. Mary cannot, for instance, know that apples are red until she experiences a red sensation.

The possible response to my argument that has the most weight is that which argues that “apples are red” is not actually a physical fact, and so while Mary may not know it in the black and white room, this does not mean that her knowledge of physical facts is incomplete. What sets facts like “apples are red” apart from physical facts is that it contains a phenomenal concept component (namely the phenomenal concept *red sensation*) and so one needs to possess this phenomenal concept (which can only be done *a posteriori* because of the nature of the experience thesis) to even understand the proposition *as a whole* let alone know it. While I do not necessarily share this view about the difference between phenomenal and physical facts, I argue that even if it were to be accepted, it would actually help provide a different (and even broader reaching) defence of physicalism. This is, that it shows that the psychophysical conditional is a phenomenal fact because it has a phenomenal component. And if this is so then it is only knowable *a posteriori*, yet it can still be true and necessary, and this also accounts for why it appears to be contingent.

I see phenomenal concepts as minimal properties of experiences that are contextualised and qualified by logic and physical facts but are also necessary for us to have an understanding of physical facts.²⁵ In this way physical facts and phenomenal concepts are intrinsically linked. It is our attempts to separate these concepts that seem to be responsible for many of the problems in philosophy of mind, and certainly for our problems with

physicalism. To understand them as essential to one another, however, as I have tried to do in this paper, provides us with strong ways of defending physicalism, as we have just concluded.

Bibliography

- Block, Ned, Owen Flanagan, and Güven Güzeldere, (eds), (1997), *The Nature of Consciousness: Philosophical Debates*, (Cambridge, MA: The MIT Press).
- Braddon-Mitchell, David, and Frank Jackson, (1996), *Philosophy of Mind and Cognition*, (Oxford: Blackwell Publishing).
- Chalmers, David, (1996), *The Conscious Mind: In Search of a Fundamental Theory*, (Oxford: Oxford University Press).
- Dennett, Daniel, (1997), “Quining Qualia”, in Block et al, *The Nature of Consciousness*, pp.619-642.
- Hill, Christopher, and Brian McLaughlin, (1999), “There are Fewer Things in Reality Than are Dreamt of in Chalmer’s Philosophy”, *Philosophy and Phenomenological Research*, vol.59(2), pp.445-454.
- Jackson, Frank, (1982), “Epiphenomenal Qualia”, *The Philosophical Quarterly*, vol. 32, pp.127-136.
– (1986), “What Mary Didn’t Know”, *The Journal of Philosophy*, vol.83, 1986, pp.291-295.
- Lewis, David, (1997), “What Experience Teaches”, in Block et al, *The Nature of Consciousness*, pp.579-596.
- Loar, Brian, (1997), “Phenomenal States”, in Block et al, *The Nature of Consciousness*, pp.597-616.
- Nida-Rümelin, Martine, (1995), “What Mary Couldn’t Know: Belief about Phenomenal States”, in W. Lycan (ed.), *Mind and Cognition: A Reader*, (Oxford: Blackwell), pp.219-241.
- Papineau, David, (1998), “Mind the Gap”, *Philosophical Perspectives*, vol.12, pp.373-388.
- Stoljar, Daniel, (2005), “Physicalism and Phenomenal Concepts”, *Mind and Language*, vol. 20(5), pp.469-494.
- Tye, Michael, (1999), “Phenomenal Consciousness: The Explanatory Gap as a Cognitive Illusion”, *Mind*, vol.108 (432), pp.705-725.

Footnotes:

- 1 This was first presented in Jackson (1982) and (1986).
- 2 These ways that the knowledge argument can be seen to respond to physicalism do not differ greatly, and being mutually consistent, they can be seen as just two ways of seeing the one problem. However, it should be noted that the ways that Mary’s example can be interpreted do differ in so far as the first way I propose shows a conclusion that appears to be pretty unique to the knowledge argument, whereas the second conclusion could be said to be shared by the conceivability argument. (For a full discussion of the conceivability argument see Chalmers, (1996)).
- 3 I take this way of expressing the psychophysical conditional from Daniel Stoljar. (Stoljar, 2005, p.472).
- 4 This is, of course, not the *only* explanation – may people argue that while the psychophysical conditional is not necessary *a priori*, it is still necessary *a posteriori*. They do, however, then have to explain why it appears to be contingent. (For more on the argument for the necessary *a posteriori* nature of the psychophysical conditional see Lewis (1997), Hill and McLaughlin (1999), Papineau (1998) and Tye (1999), and in response to these positions, Stoljar (2005)).
- 5 This is not to suggest that there is not *also* something wrong with the other premise and the conclusion, but it is the first premise that I will be focusing on here – for if it can be successfully shown that it is at fault then there is no need to respond to the other premise or conclusion (at least not in the context of this version of the knowledge argument).
- 6 We can leave out cases of non-red apples for this particular example.
- 7 Jackson, 1982, p.130.

- 8 Nida-Rümelin, 1995, p.221.
- 9 In Nida-Rümelin's example, Marianna is actually transferred to a highly coloured room, but one where the object she sees are highly controlled so that she cannot just make the logical connection that an item has a certain colour in virtue of what kind of item it is. This, however, is unnecessary for my use of the Marianna example at this point. Nida-Rümelin, 1995, pp.221-222.
- 10 Marianna's mistake also shows us that we need to have knowledge of logic and physical facts to be able to interpret and apply phenomenal concepts – to be able to give any meaning or context to these concepts. However, it is clear that this can only be done after the phenomenal concepts are possessed. You can't give clarity or context to something that you don't know to begin with.
- 11 Nida-Rümelin, 1995, p.221, footnote 5.
- 12 Jackson, 1982, p.130.
- 13 Jackson, 1986, p.291.
- 14 For such an argument see Dennett, 1997.
- 15 Papineau, 1998, (p.5) and Tye, 1999, (p.709).
- 16 Nida-Rümelin, 1995, p.221.
- 17 A strategy that has similar premises to mine, but attacks the conclusion of the knowledge argument rather than the first premise. (For arguments from the phenomenal concept strategy, or something thereabouts, see Lewis (1988), Hill and McLaughlin (1999), Papineau (1999) and Tye (1999), and in response to these positions, Stoljar (2005)).
- 18 Stoljar, 2005, pp.486-487.
- 19 Stoljar, 2005, p.487.
- 20 Stoljar, 2005, p.487.
- 21 Stoljar, 2005, p.487.
- 22 Braddon-Mitchell and Jackson, 1996, p.136.
- 23 Arguably John Searle's "Chinese room" thought experiment as well as Ned Block's two famous thought experiments "The China Brain" and "Blockhead" all attempt to show that possession of phenomenal concepts is necessary for understanding, although I do not have the space to fully work through this intuition in this paper.
- 24 This happens to be the same approach to the phenomenal concept strategy as outlined in Stoljar, 2005. Stoljar's objection to this argument is, however, that just because we cannot deduce the conclusion of a conditional from its antecedent, does not mean that it is *a posteriori*. (pp.478-479) I believe that my conclusion here stands up to this objection by showing that it is the *complete* proposition of the psychophysical conditional, as a phenomenal fact, that is *a posteriori* as opposed to just its consequent.
- 25 As we have seen in §4, the fact that possession of some phenomenal concepts is necessary for knowledge of physical facts is not the contentious issue here. So this sentence remains true, even if you believe in a strong distinction between physical facts and phenomenal facts.

Frege and Wittgenstein on Realism

Steven Bond

Limerick University

It is customary to locate the philosophy of the early Wittgenstein firmly within the mainstream of twentieth century analytic philosophy. Wittgenstein himself, in the Preface to the *Tractatus*, remarks upon his indebtedness “to Frege’s great works and to the writings of my friend Mr Bertrand Russell.” (1974 4) Without attempting to detract from the core influence of Frege on the thought of the early Wittgenstein, this paper aims to redress the traditional neglect of their philosophical differences. It is my contention that an overemphasis of the analytic inheritance of Wittgenstein has led to the fallacious assumption that the *Tractatus*, in keeping with this tradition, presents us with a realist (and furthermore an anti-idealist) worldview. Wittgenstein was equally versed in the continental tradition of Kantian metaphysics and it is from this tradition that he inherits his conceptions of philosophical realism and idealism. Unlike the analytic tradition, which conceives of realism and idealism as presenting us with an ‘either/or’ choice, the Kantian tradition presents us with a *transcendental* idealism, which is compatible with realism. This paper begins with an explication of these two varying forms of idealism and proceeds to argue for Wittgenstein’s inheritance of the Kantian rather than the Fregean strand. The latter part of the paper focuses upon the practical implication of these alternate conceptions of the realism/idealism dynamic, which finds its most lucid expression in Frege and Wittgenstein’s diametrically opposed attitudes to the nature of scientific practice.

Our starting point is Kant’s famous distinction between the phenomenal realm of spatio-temporal appearances, to which we are all privy, and the noumenal realm of the thing-in-itself, which is beyond the limits of human comprehension. For Kant, the mind does not passively receive naked perceptions. Rather, the cognitive framework of the subject actively creates, prior to perception, the spatio-temporal framework to which any such perceptions must necessarily conform. Space and time are thus granted an empirical reality i.e. they possess an objective validity with regards to all possible objects of our senses. That space and time are necessary forms of any and all potential objects of the subject’s sensible intuition, points to the fact that it is within the subject that space and time now possess this reality. Without these subjective conditions of experience, however, space and time amount to nothing – Kant terms this their transcendental ideality. It is paramount to an understanding of Kant that this idealism (transcendental or critical idealism) be separated from Berkeleyan idealism (pure idealism). The distinction arises as an insertion in Kant’s copy beside the section of the ‘Aesthetic’ concerning the transcendental ideality of space:

Pure idealism concerns the existence of things outside us. Critical idealism leaves that undecided, and asserts only that the form of their intuition is merely in us. (Kant, 2000 161)

It is the failure of others to recognize these two distinct forms of idealism that prompts Schopenhauer to claim that nothing is so consistently misunderstood as idealism, “since it is interpreted as meaning that the empirical reality of the external world is denied.” (1969 7)¹ Such a denial is the stuff of pure, or Berkeleyan, idealism. Kant, it is clear, while remaining a

transcendental idealist, does not offer us the traditional idealist rejection of the existence of objects in space. There is no rejection of realism here. As this space itself now exists only as part of my perceptual framework, the traditional problem of bridging the divide between an *internal* representation and an object existing in an *external* space simply vanishes. The crux of transcendental idealism is this bypassing of the old question of the ideality of outer appearances (i.e. whether or not things exist *out there* as we perceive them *in here*). The traditional dichotomy of realism contra idealism vanishes with the Kantian internalization of space, which makes the notion of an '*out there*' nonsensical from the offset. With Kant, one has acquired the possibility of being both an empirical realist, meaning that objects can be affirmed as existing in space, and a transcendental idealist, because the very notion of space itself now pertains only to the cognitive framework of the perceiving subject. It is this compatibilism of realism and idealism that this paper supposes is equally evident in the writings of Wittgenstein, a fact that demarcates him from the analytic tradition, whose alternate conception of realism is outlined with specific reference to Frege's '*Der Gedanke*' below.

Analytic philosophy, from its inception with Frege, is a tradition that has placed anti-psychologism high on its agenda. The inclusion of what were once held to be mind-dependent entities into a greatly expanded ontology, marked the tradition as offering a realist account of the external world. Furthermore, consequent upon Frege's inadequate knowledge of Kantian metaphysics, the analytic realism he endorses is not compatible with idealism, but rather presents idealism with the blunt pre-Kantian opposition of an antithesis. In '*Der Gedanke*', for example, Frege offers an extreme form of Platonism, in which the fundamental elements of his ontology are thoughts, conceived of as independent of any subjective consciousness. Thoughts are neither mental entities nor acquired through impressions of sense; as such, they exist neither in the external world nor in the subjective mind. Frege thus came to place his objectively existing thoughts in a third realm.²

On citing Frege's '*Der Gedanke*', it becomes clear that Frege, in his attack on idealism, has not conceived of idealism in the Kantian sense. The form of idealism he attacks is the pre-Kantian idealism that questions the very existence of the external world.

I have assumed several times that the same thing as I see can also be observed by other people. But what if everything were only a dream? If I only dreamed I was walking in the company of somebody else, if I only dreamed that my companion saw the green field as I did, if it were all only a play performed on the stage of my consciousness, it would be doubtful whether there were things of the external world at all. (Beaney, 1997 337)

Now, it is clear from the fact that Frege holds the existence of independent, external objects as open to doubt (as in the case of pure idealism), though not as a metaphysical impossibility (as with Kantian critical idealism), that the idealism against which he directs his polemic is the traditional kind. And again, when Frege is categorizing what he believes to be the distinctions between an inner idea and a thing from the outer world, it is evident that he has failed to make the Kantian leap beyond realism and idealism. For such a leap would initiate the breakdown of the distinction which Frege maintains (again taken from '*Der Gedanke*') in holding that

... ideas need an owner. Things of the outer world are on the contrary independent. (1997 334)

Frege conceives of his own empirical realism as opposed to simple idealism (though he

himself remains unaware of the distinction between simple and critical idealism). The fundamental elements of Frege's ontology are thought contents, which exist independently of the subject's consciousness. He compares the existence of a thought prior to its being grasped to a star that has existed prior to its discovery. Frege's philosophical system is thus metaphysical realist at base.³ Wittgenstein, we have said, while accepting the truth of an empirical realism, offers us the simultaneous Kantian endorsement of transcendental idealism. This finds its clearest expression in a remark made in his *Notebooks* – "...idealism leads to realism if it is strictly thought out." (1990 144) That idealism and realism can be equated points to an endorsement of the Kantian belief that transcendental idealism is but a form of empirical realism.⁴ We may add that the cryptic remark of *Tractatus* 5.64, that solipsism, when its implications are followed out strictly, coincides with pure realism, loses its opacity when considered as an affirmation of the Kantian identification of transcendental idealism and empirical realism. It is a conception of realism that is at odds with the anti-psychologism of the analytic tradition, within the frame of which Wittgenstein's *Tractatus* is typically interpreted. Wittgenstein's system is thus, like Kant's before him, metaphysical idealist at base, a point which Frege was unable to grasp.

In a correspondence with Wittgenstein regarding Wittgenstein's reaction to Frege's own paper 'Der Gedanke', Frege had the following to say.

But I would like to know what deep grounds for idealism you think I have not grasped. I take it that you yourself do not hold the idealist theory of knowledge to be true. So, I think, you recognise that there can, after all, be no deep grounds for this idealism. The grounds for it can then only be apparent grounds, not logical ones. (Monk, 1990 190)

P.T. Geach, in the preface to Frege's *Logical Investigations*, a collection of three papers of which 'Der Gedanke' is the first, remarks how interested Wittgenstein had been in the plan of Geach and Max Black to write a book of Frege translations.

He advised me to translate 'Die Verneinung', but not 'Der Gedanke': that, he considered, was an inferior work — it attacked idealism on its weak side, whereas a worthwhile criticism of idealism would attack it just where it was strongest. Wittgenstein told me he had made this point to Frege in correspondence: Frege could not understand — for him, idealism was the enemy he had long fought, and of course you attack your enemy on his weak side. (1977 vii)

We have already quoted 'Der Gedanke' to the effect that the form of idealism it attacks is a pre-Kantian pure idealism that stands in binary opposition to realism. Any attack upon pure idealism, however, to one versed in the Kantian tradition that has moved beyond the old dichotomy of pure idealism contra realism, is somewhat superfluous. And it is for this reason that Wittgenstein remains unimpressed by Frege's attack upon idealism's 'weak side'. A worthwhile critique of idealism would be concerned with the transcendental idealism of the continental tradition. It is on account of the fact that Frege presents the idealist as imposing a practical doubt on the existence of external objects, as opposed to a Kantian reduction of external objects to the status of representation, that Wittgenstein advises against the translation of Frege's 'Der Gedanke'. It is impossible for Frege, given his own typically analytic conception of idealism as the antithesis of realism, to grasp Wittgenstein's conception of the relationship between realism and idealism. Wittgenstein's conception of such is taken instead from the continental, Kantian tradition. Again, it is a tradition which places realism and idealism not in opposition to each other, but supposes instead that empirical realism is a form of transcendental idealism.

Having distinguished between two alternate conceptions of realism and idealism and having furthermore located Wittgenstein, in this regard at least, firmly within the continental camp, we come to the aforementioned practical implications that this has on the Tractarian model of science.⁵ This attitude to scientific truth is one central aspect of the divergence of Wittgenstein from Frege on account of the former's acceptance of the deep grounds for idealism. For Frege, as evidenced in a further excerpt from 'Der Gedanke' below, the scientist is a discoverer of pre-existing thoughts:

The work of science does not consist in creation, but in the discovery of true thoughts. (1997 342)

The very possibility of science is here seen to rest on the denial that thought contents could be simply ideas in people's minds. (1997 339) Wittgenstein, on the contrary, views the scientific enterprise not as the discovery of abstract, autonomous thoughts, but as the arbitrary interpretation of a given set of facts. This is given its clearest expression in *Tractatus* 6.341, in which Wittgenstein describes the possibility of viewing the same picture through variously shaped meshes. The mesh is analogous to various systems that one may use to describe the world and the form of this mesh is optional. Newtonian mechanics is one possible form of description for the world of experience. Aristotelian mechanics is another. As K. Jones puts it in *Logic and Science in the Tractatus*,

Newton and Aristotle do not say anything directly about the world, rather they recommend two different ways of thinking about the world, they tell us the precise way it is possible to describe it... What his analysis of Newtonian mechanics suggests is that the so-called laws of nature reveal no necessary order *in* the world. They are optional forms of describing the facts and do not compete with the non-conventional 'form' that runs through language and the world. The laws of science do not compete with the laws of logic. (1985 19)⁶

This 'conventionalist' view which Jones correctly attributes to the Wittgenstein of the *Tractatus* with regards to scientific theory was reaffirmed in a 1931 note in *Culture and Value*:

What a Copernicus or a Darwin really achieved was not the discovery of a true theory but of a fertile new point of view. (1980 18)

Science, for Wittgenstein, is never an accurate (or inaccurate) description of the facts of the world. It simply offers various means of interpretation for the facts given. The important point is that the world does not conform to the laws of science, but rather we create the laws of science in such a way that they conform to the world.

Russell relates how Wittgenstein, as early as 1912, had reacted to a proposed paper to be written by Russell on 'Matter'.

He admits that if there is no Matter then no one exists but himself, but he says it doesn't hurt, since physics and astronomy and all the other sciences could still be interpreted so as to be true. (Monk, 1990 47)

Here, six years prior to Frege's publication of 'Der Gedanke', Wittgenstein has already conceived of idealism such that it does not conflict with the possibility of scientific truth. Thus, idealism is already conceived such that it is consistent with realism. This is all once again related to the fact that transcendental idealism does not preclude one from an

acceptance of empirical realism. While no restriction is being placed on the quest for what may yet be termed objective truth within a clearly defined arena (the empirical), this does not alter the fact that this arena is itself subject to solipsistic restrictions at the transcendental level. As, for Wittgenstein, we are excluded entry into or knowledge of this transcendental realm, Frege is permitted his quest for scientific objectivity within the limited playing field of the empirical.

That Wittgenstein's *Tractatus* is thus *compatible* with the realism of Frege and, indeed, the entire analytic tradition, is no grounds for supposing that he enacts a similar rejection of idealism. And the dominant narrative of Wittgenstein's development, beginning as it does with Tractarian realism, perpetuates the tendency to marginalise the continental inheritance of the early Wittgenstein. And it is this obscuring of Wittgenstein's Kantian inheritance, both a deep rooted idealism and the inextricably linked conventionalist view of science, that led directly to a misunderstanding of Wittgenstein among his analytic contemporaries. We have focused in this short paper upon the case of Frege, which is but one specific instance of a more general trend.⁷ His request that Wittgenstein explain his apparent 'deep grounds for idealism' is the last recorded communication between the two. We can only suppose that it was apparent to Wittgenstein, as it seems to us now, that translating these transcendental grounds to Frege, operating as he was outside the discourse of the Kantian tradition, was something of a lost cause.

Bibliography

- Beaney, M. (ed.), 1997. *The Frege Reader*, Oxford: Blackwell Publishers.
- Burge, T., 1992. "Frege on Knowing the Third Realm." *Mind*, 101(404), 633-650.
- Engelmann, P., 1967. *Letters from Ludwig Wittgenstein with a Memoir*, Oxford: Basil Blackwell.
- Geach, P.T. (ed.), 1977. *Gottlob Frege Logical Investigations*, Oxford: Basil Blackwell.
- Hacker, P.M.S., 1986. *Insight and Illusion: Themes in the Philosophy of Wittgenstein*, Oxford: Clarendon Press.
- Hacker, P.M.S., 1996. *Wittgenstein's Place in Twentieth-Century Analytic Philosophy*, Oxford: Blackwell Publishers.
- Jones, K., 1985. "Logic and Science in the Tractatus." *Explorations in Knowledge*, 2, 18-24.
- Monk, R., 1990. *Ludwig Wittgenstein: The Duty of Genius*, London: Vintage Books.
- Moore, G.E., 1903. "The Refutation of Idealism." *Mind*, 12(48), 433-453.
- Russell, B., 1959. *My Philosophical Development*, London: George Allen & Unwin Ltd.
- Schopenhauer, A., 1969. *The World as Will and Representation*, 2nd vol., New York: Dover Publications.
- Wittgenstein, L., 1974. *Tractatus Logico-Philosophicus*, London: Routledge & Kegan Paul.
- Wittgenstein, L., 1980. *Culture and Value*, Oxford: Basil Blackwell.

Footnotes:

- 1 Schopenhauer too, who represents Wittgenstein's first encounter with philosophy, inherits the Kantian equation of transcendental idealism with empirical realism. For Schopenhauer, Locke's positive assertion of the reality of matter (materialism or realism), and Berkeley's opposing denial of matter (idealism) are equally viable alternatives, though equally blind in one eye: "It is just as true that the knower is a product of matter [realism] as that matter is a mere representation of the knower [idealism]; but it is also just as one-sided." (Schopenhauer, 1969 13) Neither Locke nor Berkeley, beginning from matter and intellect respectively, can ever cross the bridge they have created between the two. Schopenhauer is, in effect, standing on the bridge when, under the influence of Kantian transcendental idealism, he asserts of matter and intellect that "both stand and fall together; the one is only the other's reflex. They are in fact really one and the same thing, considered from two opposite points of view." (1969 16)

- 2 To quote 'Der Gedanke', So the result seems to be: thoughts are neither things in the external world nor ideas. A third realm must be recognized. Anything belonging to this realm has it in common with ideas that it cannot be perceived by the senses, but has it in common with things that it does not need an owner so as to belong to the contents of his consciousness. (1997 337)
 Kant, and Wittgenstein after him, attempted to steer a middle course between realism and idealism through the simultaneous acceptance of both. Thoughts for them are both ideas (or, more correctly, representations) and yet belong to the world, though this world is itself no longer 'external'. In this regard, Frege stands in direct contrast, as illustrated in 'Der Gedanke' above. Instead of a double acceptance of thoughts both as objects in the world and also as ideas, Frege presents us with a double rejection, leaving him only with the option of a third realm in which to house his objectively existing objects.

- 3 Russell too came to advocate realism, but was influenced on this account not by Frege but by Moore, who had led the (Cambridge) revolt against Absolute Idealism in such articles as *The Refutation of Idealism*, published in *Mind* in 1903.
 The trivial proposition which I propose to dispute is this: that *esse* is *percipi*... What I propose to show is that, in all the senses ever given to it, it is false. (1903 436)
 Russell, himself for a period an idealist, abandoned this idealism for a similarly extreme form of Platonistic realism, under direct influence of Moore. Speaking in *My Philosophical Development* (1959 48), Russell explicitly characterizes his turning towards the new logical philosophy as concomitant with a sense of freedom from his former idealism. And although Russell came to abandon such extreme Platonistic realism in 1910, he never again returned to its idealist counterpart. It is true then, not just of Frege, but of the analytic tradition in general, that it endorses a brand of realism that stands as the direct antipode of *all* forms of idealism.

- 4 This point is best captured by P.M.S. Hacker, in a footnote to the second edition of *Insight and Illusion*. The footnote appears as a direct response to David Pears' objection to Hacker's own unqualified use of the phrase 'objects of experience' in relation to Wittgenstein's *Tractatus*. Pears' objection was to state that Hacker here equivocates between taking the term 'experience' to be a reference to sense data or a reference to public objects. Hacker's retort is to state that "Wittgenstein would not have thought of himself as forced to choose between these alternatives. What is experience would be conceived, in Kantian terms, to be phenomenally real but transcendently ideal." (1986 87)

- 5 We have mentioned that Frege, in 'Der Gedanke', came to grant thoughts an objective existence in what he termed the 'third realm'. It is furthermore clear that anything belonging to this third realm is atemporal, non-spatial, and independent of causal influence. Tyler Burge, in *Frege on Knowing the Third Realm*, recounts how they are, like Plato's Forms, fundamental. In so doing, Burge draws attention to the close link that exists between Frege's rejection of idealism and his acceptance of objective scientific truth.
 Frege's Platonism shows itself in two ways. One is that he never enters the commentary that an idealist (or a deflationist) would enter on his claims about non-spatio-temporal entities, or about their objectivity or their discoverability. He takes them to be fundamental. The other is that he claims, more than once, that the assumption of the relevant entities explains the inter-subjective objectivity of science and communication. (1992 638)
 It is Wittgenstein's acceptance of idealism which accounts for his alternative, conventionalist view of science.

- 6 We are told that Newtonian mechanics imposes one form of description on the world. Thus, the

laws of mechanics do not govern things themselves, but are laws that we impose upon things in our description of the world. Jones adds that there does exist a further non-conventional 'form' in the world. As with the transcendental idealism of Kant, we find with Wittgenstein that causality, for example, is not a mere scientific convention. *Tractatus* 6.35 states that

Laws like the principle of sufficient reason, etc. are about the net and not about what the net describes. (1974 83)

Causality (which is the foundation of all our scientific worldviews) belongs to the description we impose upon the world, and not to the world itself. It is not, however, an arbitrary scientific law. It is true that we may be indoctrinated into the Newtonian worldview to such an extent that we cannot but concede its truth over, say, the worldview of Aristotelian mechanics. But we can at least conceive of the Aristotelian worldview with sense. We cannot, on the other hand, speak with sense or meaning of a worldview which is not causal in nature. Whatever shaped mesh we arbitrarily choose to view the world through, causality must remain one its non-arbitrary features.

- 7 The tendency to minimalise those aspects of Wittgenstein's early philosophy which do not coincide with staunchly analytic views is as old as the publication of the *Tractatus* itself. Russell, in his Introduction to the work, views Wittgenstein as concerned primarily with the principles of symbolism. What is not part of logical enquiry becomes, for Russell, but the application of the results of logical enquiry to traditional philosophical problems. Wittgenstein is cast as attempting to extricate us from the fields of the ethical and the mystical, which are areas of philosophical discourse now conceived as arising out of linguistic confusion. From this point of view, it becomes impossible for Russell to grasp the nature of Wittgenstein's mystical attitude, which lies beyond symbolic representation. Referring in his introduction to the *Tractatus*, Russell remarks of Wittgenstein's 'das mystische' that "His attitude upon this grows naturally out of his doctrine in pure logic." (1974 xxiii) Again, in *My Philosophical Development*, Russell, having referred to Wittgenstein's belief in a logical structure that is common to a proposition and the fact it represents, adds that "He made it the basis of a curious kind of logical mysticism." (1959 84-85) As Frege was debarred from an understanding of Wittgenstein's deep grounds for idealism, so Russell cannot, from the standpoint of an analytic empiricism, fully appreciate the significance of the mystical attitude to Wittgenstein's Tractarian worldview. This tendency to reduce those continental aspects of the early Wittgenstein (idealism, mysticism) to inessential accretions has permeated through secondary commentaries ever since Russell set them off on an analytic footing. Alexander Maslow, Eric Stenius, David Favrholt and P.M.S. Hacker are notable exceptions to this general trend. Nevertheless, much work remains to redress the balance entirely. Paul Engelmann made the following (still largely relevant) comment on the *Tractatus* as early as 1967 - "Russell's introduction may be considered one of the main reasons why the book, though recognized to this day as an event of decisive importance in the field of logic, has failed to make itself understood as a philosophical work in the wider sense." (1967 117)

Is Evolution Progressive? DIFFERENT VIEWS FROM TWO PALAEOLOGISTS*

Neralie Wise
Macquarie University

In 1543 Copernicus published *De revolutionibus orbium caelestium* in which he offered a mathematical argument for placing the sun, not the earth, closest to the centre of the universe.¹ His heliocentric view sent shock waves through the religious community because, by implication, it contradicted the belief in the anthropocentric centre of the universe. The controversy between science and religion deepened further when Charles Darwin presented his theory of evolution claiming human beings, like all other creatures, evolved from lesser organisms.² They were not created fully formed and placed in a garden. This seriously undermined the traditional religious view of humanity's significance in the over-all scheme of things. Although this debate between science and religion periodically moves to new grounds the controversy regarding the position human beings hold in the universe has not yet been resolved. The most contemporary debate concerns the progressive and non-progressive views of the nature of evolution. This became a public issue when supporters of Intelligent Design in the USA took legal action to have creationism taught in schools as part of the science curriculum. They claim that certain features of the universe and living things are best explained by an intelligent cause, not an undirected process such as natural selection, and that Intelligent Design is a science on equal footing with current scientific theories regarding the origin of life.³

Teilhard de Chardin presents a theory that constitutes a compromise between science and religion.⁴ As a palaeontologist and a Jesuit Priest no doubt he had a need to find a viewpoint that accommodated both his religious beliefs and his scientific knowledge. Chardin believes the appearance of man was the result of 'directed evolution'. He claims orthogenesis, the steady trend of evolution in a given direction over a long period of time, and a combination of chromosomic mutation were directed in such a way as to result in the creation of humanity.⁵ Chardin does not name god as the director of evolution but speaks of *nature* having a goal or direction. Nevertheless his progressive theory of evolution leaves room for the intervention of a supernatural agent making the notion that god created humanity by directing evolution compatible with his scientific account.

The long standing scientific view of evolution holds that progression is the organising principle of evolution. It generally supports one of two main thrusts of evolution; that towards increased complexity or that towards increased intelligence or consciousness.

*As soon as living beings appeared in the seas of our planet they exhibited a clear dual trend: towards complexity and diversity, the second being, in a way, complementary to the first.*⁶

*Life's steps have always been directed towards the realisation of the richest and most highly differentiated nervous system. The quantity and quality of consciousness, one may say, have always been growing throughout geological time.*⁷

Both of these views are anthropocentric in that they place human beings at the top of the evolutionary tree of life. The progressive nature of evolution is fundamental for those who wish to allow for a supernatural element in evolutionary theory because without progress there is no need for a directing agent. More recently non-progressive accounts of evolution have been proposed. In contrast to Chardin, palaeontologist Steven Jay Gould argues that life does not progress, it diversifies. In doing so he removes human beings from the top of the tree of life and his account leaves no room for supernatural intervention. His non-progressive view of evolution is anti-anthropocentric and anti-supernatural. Consequently it has been strenuously opposed by many religious groups.

I will begin my discussion of the progressive and non-progressive views of evolution with a definition of evolution. I will then give a brief account of Chardin's progressive view based on his notion of 'complexification' as the process of increasing complexity that drives evolution and that which explains human beings' progressive evolution from the simple atomic structures that emerged at the beginning of the universe. Following this I will consider Gould's non-progressive view. This will include his emphasis on the branching and 'full house' nature of evolution and both his left and right wall theories of distribution. One of the objections that might be raised by supporters of the progressive view will be addressed. My claim is that Gould's theory successfully demonstrates that there is no biological evidence that evolution is progressive. I have added a postscript for those who may be interested in alternative views. In it I will briefly note, without addressing any of the issues they might raise, the non-progressive view of zoologist Richard Dawkins and the progressive view of natural philosopher Paul Davies.

What is Evolution?

Most people are familiar with the charts that show an amoebic form followed by a fish, an amphibian, a reptile, various animals and finally man. Evolution is the theory that claims this apparent progression is caused by natural selection. During reproduction variation in gene combinations and gene mutations cause variations in the phenotype⁸ of the resulting offspring. Those offspring best suited to their local environment have a higher survival rate than their less suited siblings. They in turn will reproduce and of their offspring, those best suited to the local environment will survive and reproduce. In this way their particular characteristics will be passed on and further modified whilst the characteristics of the less successful offspring will die out. Over many generations the accumulation of very small changes will result in an organism changing its characteristics in response to the demands of its environment. Natural selection is the mechanism that drives evolution and it is the result of the interaction of three general principles, phenotypic variation, differential fitness and heritability.⁹ However natural selection as the mechanism that drives evolution is not the issue in the debate. The question is whether or not *progression* is the organizing principle of evolution. Does natural selection respond to a pull or a thrust that moves it forward in a particular direction or does it operate according to some other principle? For holders of a religious view a pull or thrust principle is necessary to allow a causal role for a super natural being and to explain why human beings are at the top of the evolutionary tree. The difficulty they have with Gould's account is that in it there is no pull or thrust; there is no forward movement and therefore no room for god.

Teilhard de Chardin

Chardin claims energy is 'corpuscled' into matter because the most primordial form of cosmic energy is granulated. These granules (photons) materialise rapidly in extremely

numerous swarms of elements, very small and very brief in their existence. When energy becomes corporeal, that is, after the emergence of material bodies (photons) he says various forces cause 'complexification', which in pre-life is individual. Matter gradually forms closed systems, becoming more complicated and centred as each element functionally superindividualises itself as incorporation takes place. More simply, the photons clumped together and became structurally complicated, thus forming a variety of closed systems that resulted in different individuals. He notes that when life with its ability to replicate occurred it resulted in populations rather than individuals and complexification became linear. He sees it as becoming linear because replication brings with it the natural selection mechanisms that drive evolution, which for him is progressive. He suggests this "very remarkable sliding" of cosmic energy into increasingly more complex corpuscular states is the result of both compression caused by gravity and of the play of large numbers of particles jostling under pressure for long periods, resulting in the most improbable combinations.¹⁰

Chardin claims plotting the rapid progress of corpuscular complexity becomes impossible once the linear progression of evolution has produced higher beings. At this stage there is a disconcerting proliferation where every trunk and branch seems to spring out at its own angle towards some different kind of success. But he notes that this proliferation does not represent a simple fanning out. It can be seen that the succession of mutations relying on one another by addition always moves in the same direction. There is a central axis of morphological progression observable in groups of related species if we study their history.¹¹ Living groups whether insects or vertebrates, studied historically show a notable advance in cerebration. In higher living beings the nervous system shows a clear tendency to gather cephalically into increasingly large ganglia.¹² Therefore Chardin argues, we can observe different levels of consciousness by estimating the cerebral perfection of the nervous system.¹³

Chardin believes evolution responds to two pulls rather than a thrust. One pulls towards the more complicated and conscious, the other towards differentiation. Life moves forward and spreads out. It progresses on a trial and error basis, advancing in exploratory waves that produce an axis of perfection sheathed by less successful attempts. (See Figure 1) For example humans head the axis of consciousness sheathed by the other less cerebrally successful primates.¹⁴ Man appeared when one zoological fibre, the human fibre, after along period of preparation succeeded in "piercing the critical surface separating the simple psychic from the reflective psychic."¹⁵ Like all progressive theories Chardin's theory is anthropocentric in that it places humans at the top of the evolutionary tree. (See Figure 2)

All appearances suggest that man, appearing on earth at the end of the Pliocene, was a sort of final or even central objective, long pursued by nature in a repeated series of rough models and approximations.¹⁶

Steven Jay Gould

Gould criticises any view that interprets evolution as a ladder proceeding in a straight line from the simple to the complex. He claims the tree of life is bushy and evolution proceeds by a complex series of branching events. The evolution of the horse in the Middle Miocene demonstrates the complexity of this process. (See Figure 3) The anthropomorphic conclusion is reached by starting with humans, drawing a line backward to the beginning of life then straightening it out. This can only be done if we fail to take the 'full house' into account. Taking the 'full house' into account means seeing the varying individuals in a population as irreducible, and avoiding viewing nature as averages representing abstract essences.

Gould argues that development from the simple to the complex is not the result of the organising principle of progression but the result of random variation operating in an environment which has a left wall. To explain this he uses the paradigm known as the 'drunkard's walk'. The drunkard is staggering along a footpath with a wall on one side and the gutter on the other. Each stagger goes in either direction with 50 percent probability. If he hits the wall he stays there until another stagger propels him away from the wall. Only one direction is open for continuous advance and he will eventually end in the gutter. (See Figure 4) Gould explains his point in this way.

In a system of linear motion structurally constrained by a wall at one end, random movement, with no preferred directionality whatever, will inevitably propel the average position away from the starting point at the wall.¹⁷

More simply, if a wall stops a randomly moving object going left it must go right. This relates to evolution in the following way. When the smallest possible organism is subject to random variations it must become more complex precisely because it cannot become any less complex. The organism's inability to become less complex acts as a left wall. (See Figure 5) Therefore the evolutionary progression from simple to complex can be explained as being inevitable without postulating progression as an organising principle.

Supporters of the progressive view might argue that this is true if random motion has no *preferred directionality* once it moves away from the wall. But natural selection demonstrates a preference for complexity and intelligence. Fossil evidence indicates that each evolutionary change produces an organism that is either more complex or more intelligent than those that went before. For example sexual reproduction is a far more complicated way of reproducing than asexual reproduction. However natural selection favours sexual reproduction because the combining of the genetic characteristics of two different parents enables the production of offspring that may be better suited to the environment than either parent. Asexual reproduction results from the splitting of a single parent into two or more offspring. Unless the genes undergo mutation the genome of the offspring will be an exact replication of the parent giving rise to a phenotype with exactly the same characteristics as the parent with exactly the same ability to survive. Life does not just randomly evolve backwards, forwards and sideways. It has a tendency to move forward. And Chardin points out, over time "either in percentage or absolute quantity, the mass of cerebralised matter has increased (and this with increasing rapidity) in the biosphere".¹⁸ Gould might have demonstrated a statistical principle regarding left walls but this has not accounted for the trend towards complexity and increased cerebration.

According to Gould the trend towards complexity and increased cerebration is an illusion resulting from ignoring the existence of right-hand walls and only taking into account the right hand tail of a right-skewed bell curve. Right-hand walls are created by natural limitations that prevent any group of organisms from further improvement. Right walls cause left-skewed distributions as more of a population approaches maximum expression and variation declines. Left walls produce right-skewed distributions because variation is only free to expand away from the wall. If we look at the right tails of two bell-curves when comparing them instead of taking the full range of variation into account, we find false evidence of a trend. Gould argues this using for his example the disappearance of the 0.400 batting average in American baseball. A player's batting average is his ratio of hits to total times at bat. A batting average of 0.400 indicates four safe hits in every ten times at bat.¹⁹ Comparing the right tail ends of an early century batting distribution curve with a current one indicates a trend towards the decline in batting skills. Early figures show the

mean batting average of 0.260 standing well away from the right wall and variation spread out widely on both sides. The few talented men near the right wall achieved 0.400 batting averages. Later in the century the mean remained the same but the best batters averaged 0.350. (See Figure 6) It appears that the average batter's skill has remained the same but the best batters are getting worse. Gould explains that this is an illusion. The batting average of 0.400 is not a 'thing'. It is a mathematical calculation. Each player has a batting average and all the player's averages are plotted on a bell curve. This produces two tails, one for the worst and one for the best performances. Bell curves can expand and contract as amounts of variation wax and wane but the tails are part of the bell curve and cannot be meaningfully detached and considered in isolation. Today general play has improved significantly. Now the unchanged average lies closer to the wall and variation in the entire system has diminished on both sides of the mean. Top hitters are trapped by the right wall and the average players are creeping up behind them. This means the top players are closer to the mean than they were in the past, resulting in a batting average of 0.350.²⁰ The reduction in the batting average in fact demonstrates an increasing excellence in play not a decline. This highlights the problems that occur when the 'full house' is not taken into account and assumptions are drawn from a comparison of 'tails'.

Emphasising the tail end of a right-skewed distribution can result in the appearance of a clear trend that is non-existent. According to Gould, basing the claim for the general progress of evolution on the right tail of life's distribution alone is absurd, because it is small and only occupied by a tiny percentage of species whose occupation of this position through time does not form an evolutionary sequence. The earliest forms of life found in fossil records are prokaryotes or bacteria and half the history of life is the tale of the bacteria. Life began with a bacterial mode lying right next to the left wall and still maintains a bacterial mode in the same position today.²¹ Successful expansion must form an increasingly right-skewed distribution as life moves away from the left wall towards greater complexity. But the resulting thin right tail end, of which humans are a part, does not represent the whole of life. (See Figure 7) Gould claims that what we see as progress in life is really "random motion away from simple beginnings, not directed impetus towards inherently advantageous complexity."²² Life does not show a general thrust towards improvement, it just produces an occasional exemplar of complexity in the only space available, while it has maintained an unvarying bacterial mode for more than three billion years.

Conclusion

Chardin has argued that life is drawn towards increased cerebration. He bases his view on the anthropocentric belief that human consciousness is nature's crowning achievement and therefore it's long stated goal. In support of his argument he notes that the mass of cerebralised matter in the biosphere has increased over time. Gould agrees the mass of cerebralised matter has increased over time. What he criticises is the extracting of some abstract measure like 'average complexity' or 'most intelligent creature', tracing the development of this back through time, labelling the discovered trend progress, then postulating that this progress must be the defining thrust of the entire evolutionary process. Using his view of life as a 'full house' and his left and right walled theories, he demonstrates that the trends underlying the progressive view are illusionary and that life does not progress, it diversifies. Further he demonstrates how life beginning at a left wall of minimal complexity and subject to random variation must result in increasing complexity. My conclusion is that Stephen Jay Gould has successfully demonstrated there is no biological evidence that progress is an essential organising principle of evolution. However this by no

means brings us to the end of the debate. As can be seen in Paul Davies' account in the postscript, fields of scientific endeavor other than biology may be used to justify both the progressive view of evolution and placing human beings at the top of the evolutionary tree. But even if one of these accounts proves to be true the progressive view does not address the original problem raised by Copernicus when he removed the earth from the centre of the universe, taking humanity with it. There is still an explanatory gap here. Placing humanity at the top of the evolutionary tree does not place humanity at the centre of the universe and there seems something odd about the notion of humanity as the crowning glory of creation living as it were, in a side street.

Postscript

Richard Dawkins

Richard Dawkins claims evolution has no aim or purpose, but his non-progressive theory of the cause of evolution differs from that of Gould.²³ His disagreement regards the third of the three principles of natural selection, heritability. Dawkins hypothesizes that the organism is not the unit of replication underlying the evolutionary process because each generation of organisms dies. It is the genes and only the genes that are replicated and passed on from generation to generation, thus creating gene lineages.²⁴ We do not inherit characteristics. We inherit genes. He claims the *organism* is created by the genes as a means of enhancing their survival and protecting the gene lineage. He argues that gene replication together with natural selection is the cause of life's diversity and the complexity of its design.²⁵ What biologists need to explain is the fit between organisms and their environment. His explanation is that natural selection prefers the organism that has a design best suited to survival in its environment. Gould and Dawkins agree that both genes and organisms play an important role in evolution but they disagree over which is the most fundamental unit.

To prove that the three principles of natural selection, phenotypic variation, differential fitness and heritability, are all that is necessary for the production of life's diversity and complexity Dawkins created a computer program simulating evolution. Briefly this involved programming for the development through reproduction of computer generated biomorphs. The development follows a 'recursive' branching rule and is modified by nine genes that affect the branching in some minor way. A +1 or -1 mutation is added to a randomly selected gene in each generation. Development results from the resulting gene value passing to the next generation through reproduction and altering the appearance of the biomorph accordingly. Natural selection is replaced with artificial selection, in this case aesthetic appeal. In twenty-nine generations the parent 'dot' in his program evolved into a distinctly insect shaped biomorph of some complexity.²⁶ (See Figure 8) This is the starting point for his argument that genes produce organisms of various designs through random mutation, natural selection chooses those designs best suited to the environment and this leads to an ever increasing diversity and complexity of life's design. By stressing that natural selection prefers good design as Dawkins does, he provides an account of intelligent design that leaves both god and progression out of the evolutionary story.

Paul Davies

Paul Davies gives a brief description of a progressive view of evolution in his book *Are We Alone?* His position is based on the following claim.

It is clear that nature has a propensity to self-organize, that is, simple physical states

tend to arrange themselves into more complex states, entirely spontaneously and without the aid of any external manipulator.²⁷

From this he proposes the hypothesis that there is a law, which he describes as more like a tendency or a trend, of increasing organized complexity operating in the universe. He sees the laws of physics as preceding matter and energy and pictures the universe as emerging out of these laws. According to his account the operation of the law of increasing complexity eventually produces a planet with life. Additional complicated sequences of self-organization produces 'mind' and ultimately the embedded consciousness of human beings.²⁸ From this he concludes that human consciousness is not the result of evolutionary chance but a fundamental emergent property of nature. It is the natural consequence of the outworking of the laws of physics.²⁹ He states that his view differs drastically from biologists who claim that consciousness is an accidental outcome of random mutation. He concedes "that the *details* of our mentality will depend on the minor and accidental specifics of evolutionary history"³⁰ but the emergence of consciousness is more or less guaranteed. In his brief account Davies does not mention the principles of natural selection but I suggest it is implied in his 'accidental specifics of evolutionary history', although obviously he gives it a much less important role and does not regard it as the cause of evolution.

The similarity between Davies' 'increasing organized complexity' and Chardin's 'complexification' is clear. They both believe there is a force underlying increasing complexity that influences the direction evolution will take. It is not a blind or random process but is dictated by *nature*. Although Davies' claim is that the propensity to self-organize occurs 'without the aid of any external manipulator' his hypothesis is readily adaptable to those who wish to introduce a supernatural force into the evolutionary process. He claims "the laws of physics are somehow already there, underpinning everything".³¹ It could be claimed that god designed the laws of physics with the express intention of creating humanity or even that the laws are a manifestation of god. Like Chardin his view is anthropocentric in that he sees reflective consciousness (humanity) as the necessary result of increasing complexity. Whereas Chardin claims evolution pulls life towards consciousness it appears Davies sees physics as pushing life towards it. Gould has demonstrated there is no biological base for progression. Whether or not a case can be found in physics is another argument.

APPENDIX I – Illustrations

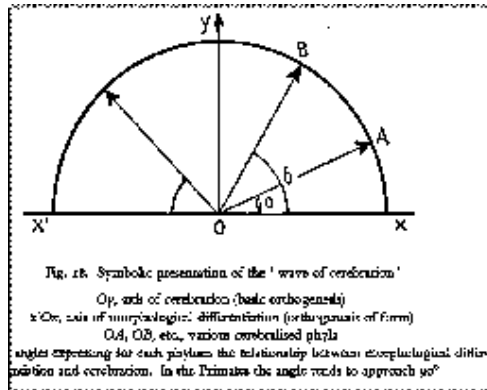


Figure 1 Symbolic Presentation of the Wave of Cerebration

Teilhard de Chardin: *The Appearance of Man* (London: Collins 1965) p 221

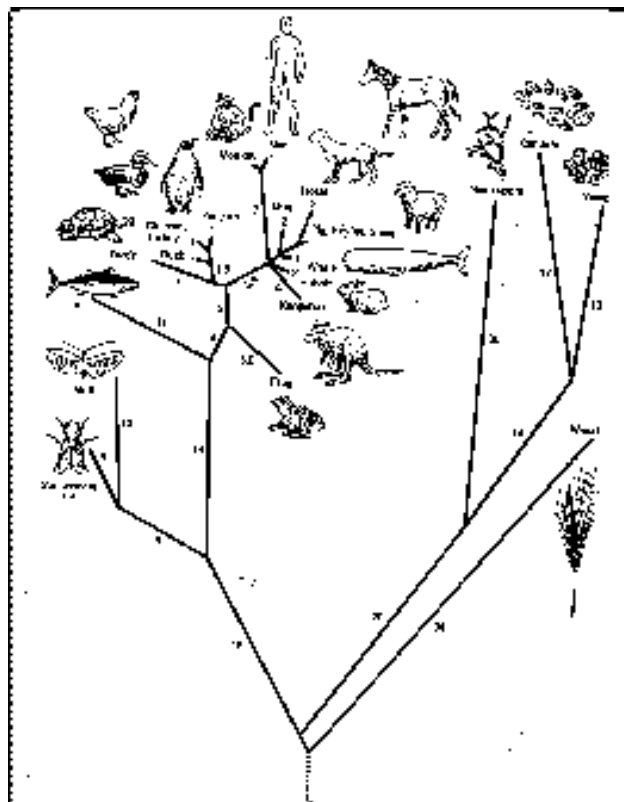


Figure 2 Anthropocentric Tree of Life

Stansfield, William: *The Science of Evolution* (New York: Macmillan Publishing Co. Inc. 1977). p 125

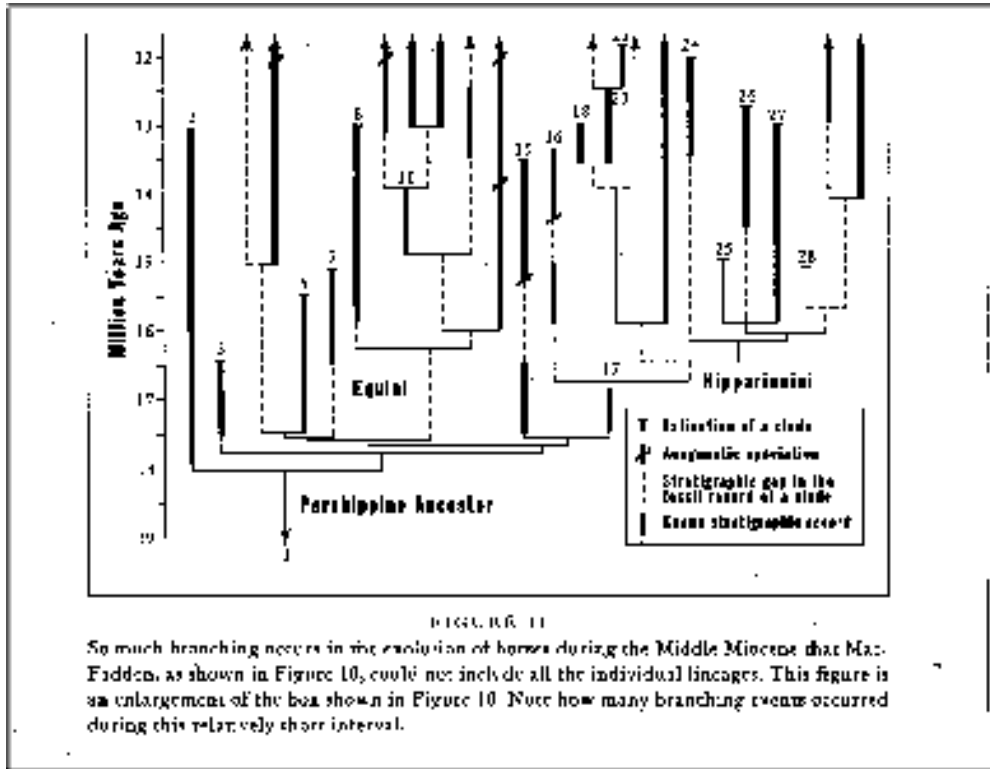


Figure 3 Branching of the Evolution of the Horse in the Middle Miocene
 Gould, Stephen Jay: *Life's Grandeur* (London: Jonathan Cape, 1996), p 66

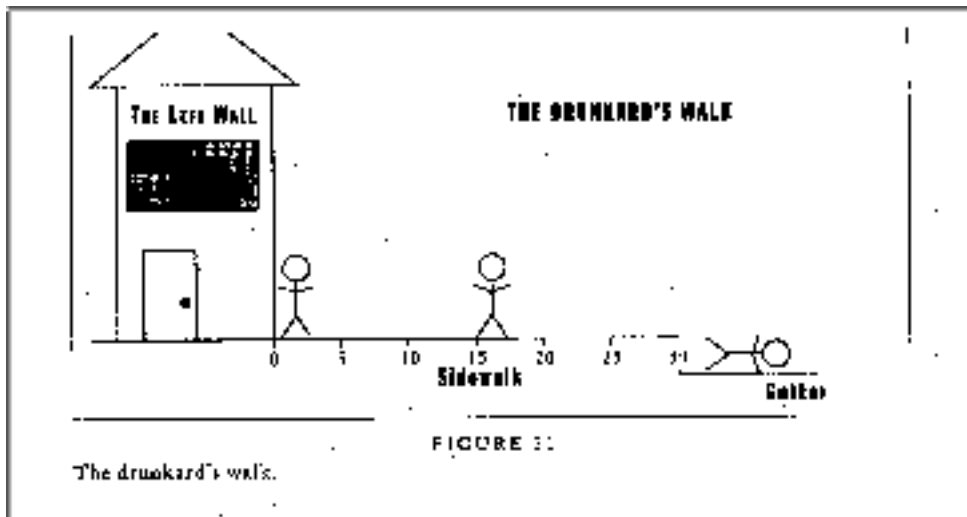


Figure 4 The Drunkard's Walk
 Gould, Stephen Jay: *Life's Grandeur* (London: Jonathan Cape, 1996), p 150

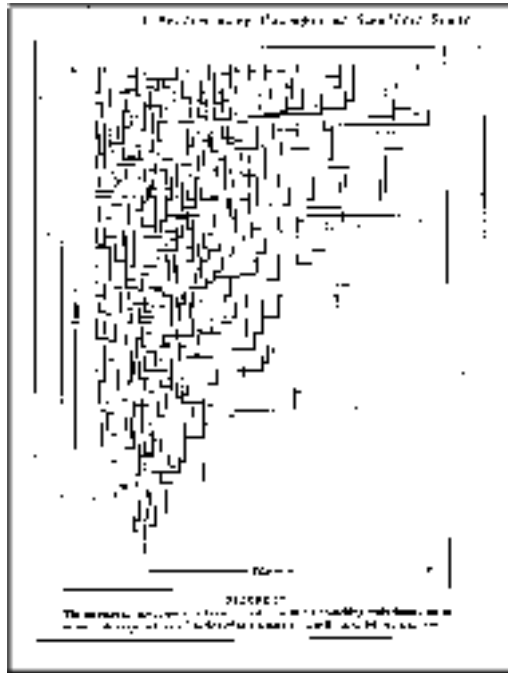


Figure 5 Branching from a Left Wall

Gould, Stephen Jay: *Life's Grandeur* (London: Jonathan Cape.1996). p 165

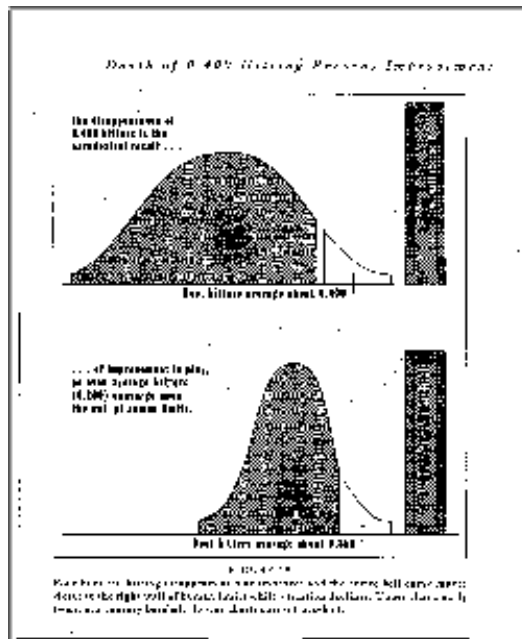


Figure 6 Death of the 0.400 Hitting

Gould, Stephen Jay: *Life's Grandeur* (London: Jonathan Cape. 1996). p 119

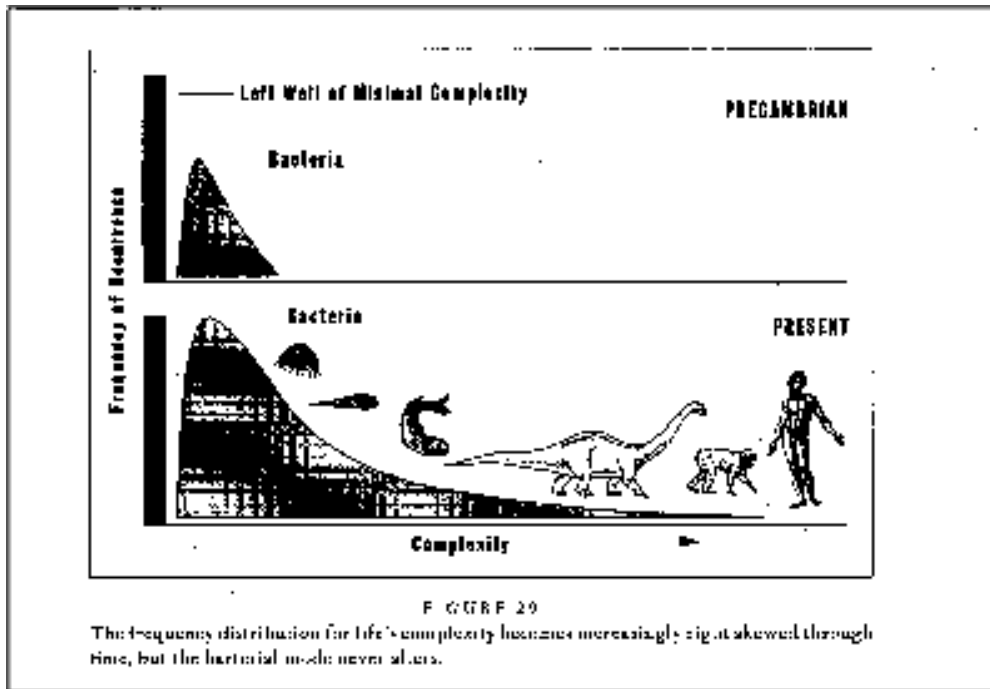


Figure 7 Right-skewed Distribution of Life's Complexity

Gould, Stephen Jay: *Life's Grandeur* (London: Jonathan Cape. 1996). p 171

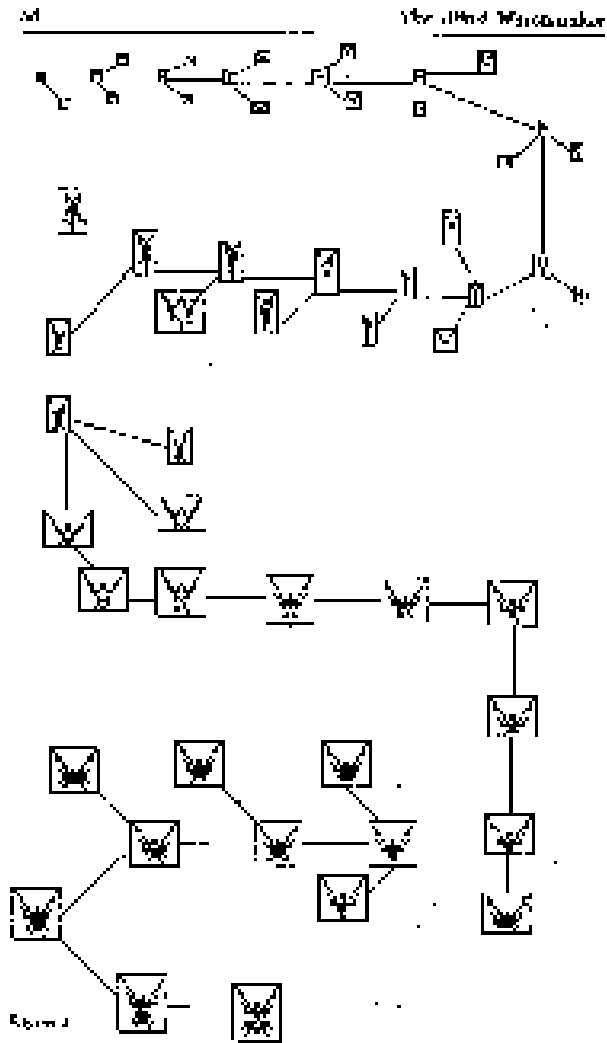


Figure 8 One Result from Dawkins' EVOLUTION' Program

Dawkins, Richard: *The Blind Watchmaker* (England: Penguin Books. 1991) p 58

Footnotes:

- * I would like to thank Dr Mitch Parsell for his suggestions and constructive comments on this paper.
- 1 Copernicus, Nicolaus: (1543) *Nicholas Copernicus On the Revolutions* (Trans.) Edward Rosen (Baltimore: The John Hopkins University Press, 1992).
- 2 Darwin, Charles: *The Decent of Man, and Selection in Relation to Sex* pt.1 (Edits.) Paul H. Barrett & R. B. Freeman (London: William Pickering, 1989).
- 3 Bowman, Kristi L.: 'Seeing Government Purpose Through the Objective Observer's Eyes: The Evolution-Intelligent Design Debates' *Harvard Journal of Law and Public Policy* 2006, pp 419-489
- 4 Teilhard de Chardin, Pierre: *The Appearance of Man* (London: Collins, 1965).
- 5 Chardin: *The Appearance of Man* p 142
- 6 Grasse, Pierre-P: *Evolution and Living Organisms* (London: Academic Press, 1977). p 10
- 7 Chardin: *The Appearance of Man* p 49
- 8 The phenotype is the sum of the characteristics manifested by an organism, as contrasted with the genotype which is the set of genes possessed by it.
- 9 Sterelny, Kim and Griffiths, Paul: *Sex and Death. An Introduction to Philosophy of Biology* (Chicago: University of Chicago Press, 1999). p 32
- 10 Chardin: *The Appearance of Man* p 214
- 11 Chardin: *The Appearance of Man* p 218
- 12 Chardin: *The Appearance of Man* p 220
- 13 Chardin: *The Appearance of Man* pp 211-217
- 14 Chardin: *The Appearance of Man* pp 218-221
- 15 Chardin: *The Appearance of Man* p 141
- 16 Chardin: *The Appearance of Man* p 34
- 17 Gould, Stephen Jay: *Life's Grandeur. The Spread of Excellence from Plato to Darwin.* (London: Jonathan Cape. 1996). p 51
- 18 Chardin: *The Appearance of Man* p 220
- 19 Gould: *Life's Grandeur* p xviii
- 20 Gould: *Life's Grandeur* pp 98-128
- 21 Gould: *Life's Grandeur* pp 170-172
- 22 Gould: *Life's Grandeur* p 173
- 23 A detailed account of the debate between Gould and Dawkins is presented in Kim Sterelny *Dawkins vs. Gould* (U.K: Icon Books, 2001). This book is temporarily out of print but is due to be re-released in the first half of 2007.
- 24 A full account of Dawkins' argument is to be found in his book *The Blind Watchmaker.* (London: Penguin, 1991).
- 25 Gould believes it is the impact of natural selection on the organism and on species that is the main driving force of evolution.
- 26 Dawkins, Richard: *The Blind Watchmaker* with appendix (London: Penguin, 1991). pp 43-74
- 27 Davies, Paul: *Are We Alone? Implications of the Discovery of Extraterrestrial Life* (London: Penguin, 1995). p 66
- 28 Davies: *Are We Alone?* p 69
- 29 Davies: *Are We Alone?* p 70
- 30 Davies: *Are We Alone?* p 70
- 31 Davies: *Are We Alone?* p 69