

# REASON PAPERS

*A Journal of Interdisciplinary Normative Studies*

## **Articles**

Debunking Popper: A Critique of Karl Popper's  
Critical Rationalism .....*Nicholas Dykes*

Is Virtue Only a Means to Happiness? An Analysis of  
Virtue and Happiness in Ayn Rand's Writings  
.....*Neera K. Badhwar*

Foundations, Philosophy, and The Location of  
Socrates' Feet .....*Nino Langiulli*

## **Discussion Notes**

Dispensing with (Men of) Reason .....*Stephen Hetherington*

Ethics and the Wittgensteinian Approach .....*Eyal Moses*

## **Review Essays**

Zoon Eleutherion: Man as an Individualist Animal  
.....*Irfan Khawaja*

Attending to Mind Itself .....*Roger E. Bissell*

**Editor:**

Tibor R. Machan

**Managing Editor:**

Mark Turiano

**Executive Editor:**

Gregory R. Johnson

**Associate Editors:**

Walter Block/ *Economics*

Douglas J. Den Uyl/ *Philosophy*

Kelly Dean Jolley/ *Philosophy*

Leonard Liggio/ *History*

Eric Mack/ *Philosophy*

John D. McCallie/ *Economics*

H. Joachim Maitre/ *International Relations*

Ralph Raico/ *History*

Lynn Scarlett/ *Political Science*

**Advisory Board:**

D. T. Armentano/ *University of Hartford*

Yale Brozen/ *University of Chicago*

Nicholas Capaldi/ *University of Tulsa*

R.L. Cunningham/ *University of San Francisco*

John Hospers/ *University of Southern California*

Isreal M. Kirzner/ *New York University*

Kenneth G. Lucey/ *SUNY College, Fredonia*

Fred D. Miller, Jr./ *Bowling Green State University*

Herbert Morris/ *University of California, Los Angeles*

Clifton Perry/ *Auburn University*

Paul Craig Roberts/ *Georgetown University*

Morton L. Schagrin/ *SUNY College, Fredonia*

Thomas S. Szasz/ *SUNY Medical Center, Syracuse*

# REASON PAPERS

No. 24 (Fall 1999)

## **Articles**

- Debunking Popper: A Critique of Karl Popper's  
Critical Rationalism .....*Nicholas Dykes* 5
- Is Virtue Only a Means to Happiness? An Analysis of Virtue and  
Happiness in Ayn Rand's Writings .....*Neera K. Badhwar* 27
- Foundations, Philosophy, and  
The Location of Socrates' Feet .....*Nino Langiulli* 45

## **Discussion Notes**

- Dispensing with (Men of) Reason ..... *Stephen Hetherington* 57
- Ethics and the Wittgensteinian Approach ..... *Eyal Moses* 73

## **Review Essays**

- Zoon Eleutherion: Man as an  
Individualist Animal ..... *Irfan Khawaja* 81
- Attending to Mind Itself ..... *Roger E. Bissell* 93

## **Reviews**

- Ann Hartle's 'Self Knowledge in the  
Age of Theory' .....*John C. McCarthy* 117
- Robert Kane's 'The Significance  
of Free Will' .....*Roger E. Bissell* 121
- Bouwsma's Notes on Wittgenstein's  
Philosophy, 1965-1975 .....*Tim Black* 131
- Stephen R.L. Clark's 'Animals and Their  
Moral Standing' ..... *Jody L. Graham* 139

# Debunking Popper: A Critique of Karl Popper's Critical Rationalism

Nicholas Dykes

## Introduction<sup>1</sup>

Karl Popper was without question one of the most eminent philosophers of the 20th Century. Author of several ground-breaking and highly influential books, and of hundreds of articles; winner of many rare prizes and other honours, such as a British knighthood; and founder of two new schools of thought, Critical Rationalism and Evolutionary Epistemology: few thinkers have made more extensive contributions to the intellectual life of their times. When he died in 1994, after a career spanning nearly 70 years, many agreed with his fellow philosophers Anthony Quinton and Rom Harré that Popper was “this century’s most important philosopher of science,” and “the last of the great logicians.”<sup>2</sup>

As the name Critical Rationalism may suggest, Popper regarded a critical attitude as the most important virtue a philosopher could possess. Indeed, he called criticism “the lifeblood of all rational thought” (PKP2 977) and, as his obituarists implied, it was towards science, and the logic of science, that his critical powers were chiefly directed. In his magnum opus, *The Open Society and its Enemies*, he wrote: “... all criticism consists in pointing out ... contradictions or discrepancies, and scientific progress consists largely in the elimination of contradictions wherever we find them. This means, however, that science proceeds on the assumption that *contradictions are impermissible and avoidable* ... once a contradiction is admitted, all science must collapse” (OSE2 39).

It is thus surprising to discover that Popper himself hardly lived up to this ideal of non contradiction. When one examines Critical Rationalism, for example, one soon notices that it is based on questionable premises; that its internal logic is seriously flawed; that it is inconsistent with other elements of Popper’s thought; and that it leads to conflicts with his own publicly stated convictions. The following pages elaborate this case.

## 1. A Brief Description of Critical Rationalism

Critical Rationalism has been referred to, by Popper himself and by others, as the theory of falsification, or falsificationism, and as fallibilism. It would be tempting, for the sake of brevity, to employ ‘fallibilism’ throughout, but the term is also associated with the founder of Pragmatism, C.S. Peirce, who actually coined it long before Popper began his career.<sup>3</sup> This paper therefore follows the lead of later Popperians such as W.W. Bartley III<sup>4</sup> and David Miller<sup>5</sup> in employing Critical Rationalism, which in any case better encompasses Popper’s thought.

The Critical Rationalism of Karl Popper (henceforth CR) begins by rejecting

induction as a scientific method. The *actual* method of science, Popper maintained, is a continuous process of conjecture and refutation: "The way in which knowledge progresses, and especially our scientific knowledge, is by unjustified (and unjustifiable) anticipations, by guesses, by tentative solutions to our problems, by *conjectures*. These conjectures are controlled by criticism; that is, by attempted *refutations*, which include severely critical tests. They may survive these tests; but they can never be positively justified: they can be established neither as certainly true nor even as 'probable'..." (C&R vii).

Elsewhere, Popper put the matter more succinctly: "all knowledge is hypothetical" (OKN 30) or "All knowledge remains... conjectural" (RASC xxxv); and it is in the form 'all knowledge is conjectural' that the essence of his philosophy has been captured - and has influenced others.<sup>6</sup>

CR was originally developed by Popper to demarcate science from non-science. He stated that for scientific knowledge to be considered *knowledge* it had to be refutable: "In so far as scientific statements refer to the world of experience, they must be refutable ... in so far as they are irrefutable, they do not refer to the world of experience" (OSE2 13).

It follows that we can never attain certainty: "The quest for certainty... is mistaken.... though we may seek for truth... we can never be quite certain that we have found it" (OSE2 375). "No particular theory may ever be regarded as absolutely certain.... No scientific theory is sacrosanct..." (OKN 360). "Precision and certainty are false ideals. They are impossible to attain and therefore dangerously misleading..." (UNQ 24). He summed up with an oft-repeated aphorism: "*We never know what we are talking about*" (UNQ 27).

Accordingly, Popper refused to grant any philosophical value to definitions: "Definitions do *not* play any very important part in science.... Our 'scientific knowledge'... remains entirely unaffected if we eliminate all definitions" (OSE2 14). "Definitions never give any factual knowledge about 'nature' or about the 'nature of things'" (C&R 20-21). "Definitions.... are never really needed, and rarely of any use" (RASC xxxvi).

Although he held these positions all his working life, Popper did acknowledge that they were open to criticism: "nothing is exempt from criticism ... not even this principle of the critical method itself" (OSE2 379).

## 2. The First Premise of CR

Popper built his philosophy on foundations borrowed from Hume and Kant. His first premise was wholehearted acceptance of Hume's attack on induction. The second, to be addressed in the next section, was agreement with Kant's view that it is our ideas which give form to reality, not reality which gives form to our ideas.

Hume, whom Popper called "one of the most rational minds of all ages" (PKP2 1019), is renowned for elaborating the 'problem of induction' - a supposedly logical proof that generalisations from observation are invalid. Most later philosophers have accepted Hume's arguments, and libraries have been filled with attempts to solve his 'problem.'

Popper thought he had the answer. "I believed I had solved the problem of induction by the simple discovery that induction by repetition did not exist"

(UNQ 52; cf OKN 1ff & PKP2 1115). What really took place, according to Popper, was CR, knowledge advancing by means of conjecture and refutation: "... in my view here is no such thing as induction" (LSCD 40); "what characterises the empirical method is its manner of exposing to falsification, in every conceivable way, the system to be tested" (LSCD 42).

Hume, said Popper, had shown that: "there is no argument of reason which permits an inference from one case to another... and I completely agree" (OKN 96). Elsewhere he referred to induction as "a myth" which had been "exploded" by Hume (UNQ 80). He further asserted that "There is no rule of inductive inference - inference leading to theories or universal laws - ever proposed which can be taken seriously even for a minute" (UNQ 146-7; see also RASC 31).

### The Problem with 'The Problem'

Popper's solution was certainly correct in one respect. The problem of induction would indeed vanish if there were no such thing as induction. However, the issue would be resolved much more positively were it to turn out that Hume had been *wrong*, and that there never had been any problem with induction in the first place. And, in point of fact, this is the case. Despite his great skill as a thinker and writer, Hume missed the point. Induction does not depend for its validity on observation, but on the Law of Identity.

Hume stated, in essence, that since all ideas are derived from experience we cannot have any valid ideas about future events - which have yet to be experienced. He therefore denied that the past can give us any information about the future. He further denied that there is any necessary connection between cause and effect. We experience only repeated instances, we cannot experience any "power" that actually causes events to take place. Events are entirely "loose and separate.... *conjoined* but never *connected*."

According to Hume, then, one has no guarantee that the hawthorn in an English hedge will not bear grapes next autumn, nor that the thistles in a nearby field won't produce figs. The expectation that the thorn will produce red berries, and the thistles purple flowers, is merely the result of "regular conjunction" which induces an "inference of the understanding."<sup>8</sup> In Hume's view, there is no such thing as objective *identity*, there is only subjective "custom" or "habit."

However, Hume also wrote: "When any opinion leads to absurdities, it is certainly false"<sup>9</sup> and the idea that one might gather grapes of thorns or figs of thistles is surely absurd enough to qualify. And false is what Hume's opinions most certainly are. Left standing, they lead to what he himself called "the flattest of all contradictions, viz. that it is possible for the same thing both to be and not to be."<sup>10</sup>

The crux of the case against Hume was stated in 1916 by H.W.B. Joseph in *An Introduction to Logic*: "A thing, to be at all, must be something, and can only be what it is. To assert a causal connexion between *a* and *x* implies that *a* acts as it does because it is what it is; because, in fact, it is *a*. So long therefore as it is *a*, it must act thus; and to assert that it may act otherwise on a subsequent occasion is to assert that what is *a* is something else than the *a* which it is declared to be."<sup>11</sup> Hume's whole argument - persuasive though it may be - is,

to borrow Joseph's words, "in flat conflict with the Law of Identity."<sup>12</sup>

Existence *implies* identity. It is not possible to exist without being *something*, and a thing can only be what it is: A is A. Any *actions* of that thing form part of its identity: "the way in which it acts must be regarded as a partial expression of what it is."<sup>13</sup> Thus to deny any connection between a thing, its actions, and their consequences, is to assert that the thing is not what it is; it is to defy the Law of Identity.

It is not necessary to prolong this discussion. Entities exist. They possess identity. By careful observation - free from preconception - we are able to discover the identities of the entities we observe. Thereafter, we are fully entitled to assume that like entities will cause like events, the form of inference we call induction. And, because it rests on the axiom of the Law of Identity, correct induction - free from contradiction - is a valid route to knowledge. The first premise of CR is therefore false.

There is nonetheless a substantial grain of truth in Hume's position, or few philosophers would have followed him. The grain lies in the *precision* of our knowledge of future events. Hume denied all knowledge of the future because we can have no experience of it. As we have seen, this is not true, it overlooks the Law of Identity. What *is* true, is that our prediction of events is limited by the unforeseeable. An 'O' ring may fail and destroy an otherwise reliable spacecraft; an icy road surface may cause a pristine Rolls-Royce to crash. For, no matter how sound our judgment or wide our experience, we cannot possibly have complete, certain and absolute knowledge of future events. We are not omniscient: all kinds of unforeseen happenings *may intervene* to spoil even the best laid of our plans, and of course new information about old subjects continuously comes to light. But, armed with the Law of Identity, there is no reason to allow the unforeseeable to turn us into sceptics. The universe is not a series of "loose and separate events" any more than time is a series of discrete, unrelated segments of duration.

It should also be noted that, in fact, all knowledge of entities, and all knowledge of language, is acquired inductively. A child's knowledge of apples, for example, is based on a very limited sampling. A student's knowledge of the word 'inference' is founded on a similarly narrow acquaintance. If it were true that induction is a myth, then all knowledge of external reality, all language, and all human thought - which depends on knowledge of reality and on language - would be myths as well, including, of course, CR.

### 3. Popper's Kantian Premise

Popper described himself as an "unorthodox Kantian" (UNQ 82); i.e., he accepted part of Kant's epistemology, but not all of it: "Kant was right that it is our intellect which imposes its laws - its ideas, its rules - upon the inarticulate mass of our 'sensations' and thereby brings order to them. Where he was wrong is that he did not see that we rarely succeed with our imposition" (OKN 68n31; cf OKN 328, C&R 48-9).

Popper's Kantianism reveals itself most clearly in his view of our senses, which he saw as creative modifiers of incoming data, not as neutral 'windows on the world': "Classical epistemology which takes our sense perceptions as

'given', as the 'data' from which our theories have to be constructed by some process of induction, can only be described as pre-Darwinian. It fails to take account of the fact that the alleged data are ... adaptive reactions, and therefore interpretations which incorporate theories and prejudices and which, like theories, are impregnated with conjectural expectations... there can be no pure perception, no pure datum..." (OKN 145).<sup>14</sup>

### **A Fundamental Difficulty**

Popper's Kantian premise raises enough issues for a book. In this short paper, there is room only for a single objection. Namely, if it is true that our senses are pre-programmed; if it is true that "*there is no sense organ in which anticipatory theories are not genetically incorporated*" (OKN 72); then what flows into our minds is *determined* and what flows out of them is *subjective*. If our senses are not neutral, if they organise incoming data using pre-set theories built into them by evolution, then they do not provide us with unalloyed *information*, but only with *prescriptions*, the content of which is determined by our genetic make up. Whatever is thereafter produced inside our heads - cut off as it is from any objective contact with reality - must be subjective.

Popper's Kantian premise thus deprives CR of universality. Since it is ultimately the product of the pre-programmed interpretation of the data which entered Popper's mind, CR is a theory which can only be applied to Popper. According to his own view of his contact with reality, he would not be able to verify the relevance of CR to anybody else.

Solipsism looms, yes, but that is a natural consequence of all theories of determinism. For if thought, or the basis of thought, is determined; whether by social class, or the subconscious, or whatever determinant is preferred; then the deterministic theory itself must be determined, according to the theory, and can only be relevant to the person who expounds it. Everybody else is determined by *their* class, subconscious, genes, material substrate, environment, or whatever it is that is supposed to do the determining. All theories of determinism are, to use Brand Blanshard's term, 'self-stultifying'.<sup>15</sup>

The objection is analogous to the one raised by Anthony Flew against those philosophers - eg Hume and Kant - who claim that we can only have knowledge of our own sense impressions. If sense data are all we can know, solipsism is the inevitable result: "mental images ... are (necessarily) private ... and (logically) cannot be accessible to public observation."<sup>16</sup>

### **Objectivity**

In *Unended Quest* Popper observed bluntly that "there is no such thing as an unprejudiced observation" (UNQ 51). Although this appears to rule out the possibility of objectivity, that was not Popper's intention. Rather, again following Kant perhaps, he thought the basis for objectivity lay elsewhere: "the *objectivity* of scientific statements lies in the fact that they can be *inter-subjectively tested*" (LSCD 44). He later restated this slightly differently: "it is the public character of science... which preserves the objectivity of science" (POH 155-6).

Unfortunately, these assertions do not bear the weight placed upon them. For if Popper's Kantian premise is true (i.e., if anticipatory theories *are* geneti-

cally incorporated into our sense organs and, therefore, there is no such thing as an unprejudiced observation) then senses would not cease to be prejudiced merely by being multiplied. The defective logic could hardly be more clear. One cannot offer as an universal affirmative proposition 'all human senses are prejudiced, i.e. subjective' then ask one's readers to accept that pooling the senses of many persons yields objectivity. If senses are subjective individually they are subjective collectively.<sup>17</sup>

To conclude under this head, it is plain - even after only a very brief treatment - that Popper's Kantian premise, far from providing CR with a secure footing, leads instead to insuperable problems, not least of which are conflicts with Popper's own rejection of determinism and subjectivism in such works as *The Poverty of Historicism* and *The Open Universe*.

#### 4. Language Difficulties

Popper called conjecture and refutation a "new way of knowing" (OSE2 383). However, from a commonsense point of view, it can immediately be objected that we do not normally claim to 'know' something which is neither "positively justified," "certainly true," nor even "probable" (C&R vii). Knowledge, for most people — and for most scientists — is something which it is possible to be sure of, to justify, to validate, to prove; in other words, *to know*.

Conjecture, on the other hand, is by definition not knowledge. According to *Chambers English Dictionary*, a conjecture is "an opinion formed on slight or defective evidence or none: an opinion without proof: a guess". Since one cannot define an idea by means of other ideas which are contrary to it, it is clearly illegitimate to place knowledge in the same category as conjecture. More pointedly, the proposition "all knowledge remains conjectural" is a contradiction in terms.

The objection gathers strength when one notices that Popper's proposition is itself *not* conjectural. Universal and affirmative, it states that "All knowledge remains conjectural" - which is a claim to knowledge. The proposition thus asserts what it denies and is self-contradictory on a second count.

Another immediate problem is that the notion of 'conjecture' depends for its intelligibility upon the prior concept of 'knowledge.' The idea of a 'conjecture' arose precisely to designate a form of mental activity which was *unlike* knowledge, and to distinguish clearly *from* knowledge an idea put forward as opinion without proof. In the Objectivist philosophy of Ayn Rand this error is known as 'the fallacy of the stolen concept.' A classic example was Proudhon's claim that 'property is theft.' But the concept of 'theft' depends on the prior concept of 'property' and would be unintelligible without it.<sup>19</sup> In exactly the same way, and to repeat, the concept of 'conjecture' cannot be understood apart from the prior concept of knowledge - *from which* it is to be distinguished. For example, 'Northern Dancer might win the Kentucky Derby' was once a conjecture. When the horse did come first, its win became an item of knowledge.

The invalidity of the proposition 'All knowledge remains conjectural' becomes even more apparent when one considers that Popper employed a large vocabulary of English and German words all of which he had to learn, and

to *know*, in order to express any or all of his ideas. There is little conjectural about the words of a language: either the German word *Forschung* means 'scientific discovery' or it does not. Similarly, in all his philosophical and scientific work Popper depended on a broad range of core concepts - evolution, energy, light, atom, mass, force, etc - all of which are normally recognised as unalterable brute facts, not as conjectures. 'All knowledge is conjectural' may sound intriguing, but throughout his career Popper actually worked within a framework of knowledge, not of conjecture.

A further problem arises when one considers the concept of 'growth' in Popper's claim that knowledge grows through conjectures and refutations. (The subtitle of his book by that name is *The Growth of Scientific Knowledge*.) A legitimate response to this assertion is: 'What exactly is it that grows?' The concept of *growth* implies the existence of a thing, a body, an entity of some sort, *that which grows*. It may well be true that conjectures and refutations play a role in the growth of knowledge, but they could hardly do this without some knowledge to work on. The growth of knowledge via conjecture and refutation presupposes pre-existing *knowledge*, not pre-existing conjectures.<sup>20</sup>

That the growth of knowledge implies *knowledge* is another illustration of Popper's dependence on something he attempted to deny, effectively 'stealing' a concept. CR is supposed to replace our commonsense idea of inductively-acquired knowledge with a more accurate one of a continuous process of conjecture and refutation. But that process would be meaningless without something for the process *to process*, and that something is knowledge, not conjecture.

Lastly, the proposition 'all knowledge is conjectural' is simply not true. The writer's observation that 'the sun is shining' is not conjectural, it is a fact known to him and countless other observers. At 2.15pm on 15 March 1999 in western England the sun is shining. The observation is no more conjectural than 'Bill Clinton is President of the USA (at time of writing),' or 'Einstein's grandparents are dead,' or 'the French for 'yes' is *oui*, or '2 plus 2 = 4.' These statements are *true*. They are demonstrable to any sane person; either ostensibly, or through the presentation of evidence beyond reasonable doubt, via simple common sense, or by means of logic. They constitute knowledge, *not* conjecture.

## 5. Problems in Practice

Other problems surface when one considers actually employing conjecture and refutation; i.e., when one looks at CR in practice. Briefly stated, the method urges us to conjecture, then to subject the resultant theory to severely critical tests. If it survives those tests, we are permitted to grant the theory a degree of verisimilitude, the more stringent the tests, the higher the degree.

The first problem is the method's apparent arbitrariness. The conjecture or theory to be tested - and Popper said the bolder the better - would presumably be selected by the tester. But no criterion for selection is given. We might be referred to an earlier CR exercise, but since that route risks infinite regress (via earlier and earlier CR exercises), the conjecture to be tested must fall outside the scope of CR. Therefore, unless further information is provided, it is not obvious how the charge of arbitrariness can be resisted.<sup>21</sup>

Consequently, the whole approach smacks of straw men. If a conjecture survives all CR tests, it could merely be that a 'virtuous straw man' (the conjecture) has one by one fended off an army of lesser straw men (the tests). But nothing would be proven by all this. Not only do we still require evidence of the worthwhileness of the conjecture, some other method is needed to show that the opposing arguments are truly exhaustive and not just straw. To use an analogy: it is perfectly possible for a dangerous lunatic to pass a driving test. Even the most stringent 'advanced driver' courses ever devised may not uncover the explosive unroadworthiness of 'the nut behind the wheel.'

The method of conjecture and refutation also appears to be a form of question begging. It must surely assume some measure of truth in the conjecture under examination, or there would be little point in the exercise. Put simply, the method states: 'My proposition deserves examination. Nothing in the process of examination undermined my proposition. Ergo my proposition has verisimilitude.' It may well have, but the proposition's *soundness* has not been established by that reasoning. One recalls the famously circular Ontological Argument for the existence of God: 'God is that than which nothing greater can be conceived. If 'that than which' didn't exist, it couldn't be 'the greatest'. Therefore God exists.' But the argument assumes in its first premise that which it sets out to establish and is clearly invalid.

The fact of the matter is that the truth of a proposition rests on the correct identification of the referents and relationships involved, not on any prior or subsequent *argumentation*. In any design, philosophical or practical, if a false identification is incorporated, whole libraries of arguments may not reveal the consequent flaws. A building can be the most beautiful ever built, but a single misplaced decimal point in a stress calculation can bring it crashing down. As Popper so rightly said: "*contradictions are impermissible and avoidable... once a contradiction is admitted, all science must collapse*" (OSE2 39).

## 6. Refutability as a Criterion of Demarcation

CR claims to distinguish science from non-science by the refutability of scientific theories. Popper's standard example was Newtonian physics, so radically displaced by Einstein.<sup>22</sup> On the other hand, Popper maintained, there were theories such as those of Marx and Freud, which were non-science because irrefutable. This was Popper's famous 'criterion of demarcation,' which he developed as a young man and held to all his life.

Relatively few philosophers have embraced it however. Tom Settle, a major contributor to P.A. Schilpp's massive festschrift, *The Philosophy of Karl Popper*, stated firmly in 1970: "As a criterion of demarcation between science and non-science, Popper's 'falsifiability'-plus-a-critical-policy does not work" (PKP2 719). Other contributors evidently agreed; among them A.J. Ayer, William C. Kneale, Imre Lakatos, Grover Maxwell and Hilary Putnam.

One can understand the importance of the distinction to the young Popper. Fascinated by science, he was surrounded by true-believing Marxists and Freudians all of whom claimed science on their side while espousing doctrines which seemed to Popper obviously false. Nonetheless, 'refutability' seems to miss the mark. The ideas of Marx or Freud stand or fall on their conformity to

logic and the available evidence - in exactly the same way as the ideas of Newton or Einstein. Marxism and Freudianism failed to survive as viable theories due to myopic concentration on a narrow range of data, false interpretations of evidence, and logical inconsistency. They never were 'irrefutable.' They failed precisely because they could be, and were, *refuted*; either by contrary evidence, by exposure of contradictions, or by the resolute refusal of reality to conform to their predictions. It wasn't *refutability* which made them unscientific, it was inaccuracy and/or illogicality.

Science is distinguished by its strict adherence to physical evidence. Non-science, on the other hand, is invariably characterised by preconception, followed by a cavalier disregard for, or rationalisation of, anything that doesn't fit into the preconceived schema. In one sense, this is what Popper was saying. But, due perhaps to his dislike of definitions, he homed in on the wrong identifying characteristic.

There are other, more serious, criticisms of Popper's theory of demarcation. Grover Maxwell pointed out that 'All men are mortal' is a perfectly sound scientific statement which is not falsifiable (PKP1 292). Popper defended himself robustly (PKP2 1037ff), but Maxwell seemed to have the stronger case. Maxwell might also have taxed Popper about mathematics. The axioms of mathematics cannot be refuted. According to the demarcation theory, therefore, mathematics is not a science. But physics is inseparable from mathematics. Quantum mechanics, for example, could hardly be expressed without it. So physics cannot be a science either. Much the same could be said about logic. The Law of Contradiction, etc, cannot be refuted, so logic is not a science.

There is besides the singularly Popperian problem of Marxism. Marxism was one of the theories which led Popper to develop his conception of demarcation in the first place: "I had been shocked by the fact that the Marxists... were able to interpret any conceivable event as a verification of their theories" (UNQ 41-2). Yet in "Replies to my Critics" Popper changed his tune: "Marxism was once a scientific theory"; "Marxism was once a science" (PKP2 984-5). No doubt Popper would have swamped this objection with distinctions between Marx and Marxism, but the notion that Marxism could both be and not be a science does little to inspire confidence in Popper's theory of demarcation.

## 7. Popper's Via Negativa

One of the most troubling aspects of Popper's philosophy is his devout refusal to consider anything positive, a negativity which reminds one of the *via negativa* of medieval theology.<sup>24</sup> The scholastic principle, "we cannot know what God is, but rather what He is not" is remarkably similar to Popper's assertion that "natural laws.... do not assert that something exists or is the case; they deny it" (LSCD 69). CR is invariably concerned with what is *not*, never with what is. Yet the negative 'it is not' cannot be uttered without implying the positive 'it.' A negative *implies* a positive, unless one is actually denying the existence of an entity, but that is a different issue.

That negative implies positive was clearly understood by Popper. He referred to "the notion of falsity - that is, of untruth - and thus, by implication,

the notion of truth" (UNQ 98). But he did not seem to see that truth implies a '*what is*' question every time CR tells us *what is not*. It is a stolen concept situation: the idea of 'falsity' depends upon the logically prior idea of 'truth.' Or, as Anthony O'Hear has expressed it: "there can, in fact, be no falsification without a background of accepted truth."

Grover Maxwell also noted this problem. He pointed out that many theories are in fact *positively confirmed* (PKP1 292ff). Yet Popper continued to insist in "Replies to My Critics" that, "we certainly are not justified in reasoning from an instance to the truth of the corresponding law... *we are* justified in reasoning from a counterinstance to the *falsity* of the corresponding universal law" (PKP2 1020).

However, recalling Popper's Kantian premise, one might reasonably enquire at this point: If all observations are theory-laden, and thereby suspect, what justifies our placing any confidence in negative observations? The procedure of observation is identical whether one is seeking evidence in favour of a theory, or testing for evidence against it. If our senses are automatically suspect, as Popper maintained, negative or falsifying instances deserve no more credibility than positive or confirming ones.

Further, remembering Popper's Humian premise, one immediately wants to ask: If we are not allowed to argue from positive instances to true laws, why are we allowed to argue from counterinstances to negative laws (we were told above that "natural laws... deny"). The *reasoning process* is the same. Collecting disconfirmations and arguing negatively scarcely differs from collecting confirmations and arguing positively. Both are inductive procedures and, as such, have been disallowed in advance by Popper's rejection of induction.

Certainly, a single negative instance suffices to refute any universal proposition. Australian black swans falsified the belief that all adult swans were white. Popper was perfectly correct to remind us of this, and also that one or more positive instances do not necessarily *establish* universal propositions. But colour never was the defining characteristic of swans. The discovery of black ones did not entitle Popper to assert that their *essential* features - long necks, powerful wings, etc - were equally suspect.<sup>26</sup>

The bottom line which CR must confront, however, is that one cannot falsify a scientific theory without inference from observed instances. However much Popper may have rejected induction, his own method was in fact dependent upon it.<sup>27</sup>

## 8. Truth, Facts and Realism

As a metaphysical realist, Popper upheld the correspondence theory of truth: "*A statement is true if and only if it corresponds to the facts*" (OKN 46). Although he reiterated this frequently (eg OSE2 369ff, UNQ 140ff), only once did he go into detail about what he meant by 'fact.' "Facts are something like a common product of language and reality... they are reality pinned down by descriptive statements.... New linguistic means not only help us to describe new kinds of facts; in a way, they even create new kinds of facts. In a certain sense, these facts obviously existed before the new means were created.... But in another sense we might say that these facts do not exist *as facts* before they

are singled out from the continuum of events and pinned down by statements - the theories which describe them" (C&R 214).

Unfortunately, neither the lines quoted, nor the rest of the passage in the book, clarify the meaning of the word 'fact.' Since Popper's claim that 'truth means correspondence to the facts' cannot be evaluated without such clarification, we turn again to *Chambers Dictionary*, which defines 'fact' as "reality, a real state of things, as distinguished from a mere statement or belief." But if this definition is correct, it leads immediately to another problem with CR.

CR states that for knowledge to be regarded as scientific it must be falsifiable. Plainly then, if an item of 'knowledge' is falsified, it can no longer be regarded as a fact. In Popper's own words, a false conjecture "contradicts some real state of affairs;" "falsifications... indicate the points where we have touched reality" (C&R 116). What we are left with are conjectures which have not yet been falsified. But a yet-to-be-falsified conjecture can hardly be called a *fact*, 'a real state of things.' It is rather 'a mere statement or belief' *from which* facts are to be distinguished.

Remembering that we have been forbidden to regard as certain anything which we may think we know about facts, all knowledge is conjectural; and that our senses are suspect because 'theory impregnated;' we are led to the seemingly inevitable conclusion that we can never know any facts. All we can 'know' are falsifiable conjectures which, as we have just seen, are not facts. Further, if this is the case, we can never find out what is true. For if truth means correspondence with the facts, as Popper assured us it did, and we cannot know any facts, then we cannot know any truth.

It could be argued that this is precisely Popper's whole philosophy. That might be correct. But so arguing would not remove the incompatibility between Critical Rationalism and Popper's espousal of the correspondence theory of truth.

It would also appear that CR conflicts with another foundation of Popper's thought, his realism. "Denying realism" he stated, "amounts to megalomania (the most widespread occupational disease of the professional philosopher)" (OKN 41). He himself had always been: "a commonsense realist.... I was interested in the real world, in the cosmos, and I was thoroughly opposed to every idealism..." (OKN 322-3). A few pages later he wrote: "whether our man-made theories are true or not depends upon the real facts; real facts, which are, with very few exceptions, emphatically not man-made. Our man-made *theories* may clash with these real *facts*, and so, in our search for truth, we may have to adjust our theories or to give them up" (OKN 328-9).

One must agree with these sentiments. But, if the arguments just outlined are correct, it is CR which is in need of adjustment. For if CR does deny us any knowledge of real facts, the theory not only contradicts realism, it leaves one with no good reason to be a realist. Secondly, if the reasoning in other sections of this essay is correct, then CR conflicts with the fact that, having discovered such real facts as the existence of the works of Karl Popper, say, we *can and do* have true knowledge of reality. No matter which way one looks at it, CR seems out of place in the mind of anyone who aspires to be a realist.

### 9. Definition and Contradiction

Popper's espousal of the correspondence theory also conflicts with his scorn for definitions. When we assert that a statement corresponds to the facts we mean that the words used accurately describe a specific external situation. But we could not assert correspondence if the words did not have precise meanings; i.e. did not have precise definitions.

Popper liked to aver, provocatively, that we never know what we are talking about. But if his aphorism were *true*, a statement such as 'arsenic is poisonous' would be vacuous. Yet arsenic does exist. It *is* a chemical substance which, ingested above a certain concentration, is very likely to kill a human being. Which means, arsenic is poisonous. The statement is true, it corresponds to the facts. But it is only true because the words employed are accurately defined.

The correspondence theory of truth refers to human *ideas*. Whether one calls those ideas 'concepts,' 'statements,' 'propositions' or 'theories,' we are only able to hold them in consciousness, to relate them to facts, and to communicate them, via the medium of words. Words are the audio-visual symbols of our ideas. In a very real sense they link us to reality. Which means that if their definitions are vague or shifting, we cannot hope to arrive at any reliable truth: no definitions, no correspondence theory. As Aristotle said: "not to have *one* meaning is to have *no* meaning, and if words have no meaning, our reasoning with one another, and indeed with ourselves, has been annihilated."<sup>28</sup>

Even more serious is the matter of contradictions. Although he held contradictions to be "*impermissible and avoidable*" (OSE2 39) Popper had previously dismissed the Laws of Thought (which of course include the Law of Contradiction) as "psychologism" and "a thing of the past" (LSCD 98). Whatever the merit of that judgment, it is difficult to see how we can uncover contradictions if definitions "never give any factual knowledge about 'nature' or about the 'nature of things'" (C&R 20-21) which statement must imply that there is no significant connection between words and facts. Indeed, it is hard to see how logic and the Law of Contradiction are *possible* if discussions of the meaning of words - i.e., of their relationship to facts - are "tiresome phantoms" or "verbal quibbles" as Popper insisted (eg C&R 28, or TOU xxi). The upshot here is that the Law of Contradiction, far from being all-important to science, as Popper so vigorously implied, seems *excluded* by CR. If all identifications are conjectural, just 'guesses,' and definitions of no value, we would not be able to identify subject and attribute positively enough to show that they do, or do not, belong together.

### 10. Popper's Three World Theory

Early in his career, Popper began developing a theory in which he split reality into three parts: the physical world, or the world of facts; the world of consciousness, of mental processes and events; and a third world, the products of the human mind, which he called 'objective knowledge.' Popper obviously regarded the theory as important and described it in detail several times (eg OKN 106ff, & 152ff). The following is from his autobiography, *Unended Quest*:

"If we call the world of... physical objects... the *first world*, and the world of

subjective experiences... the *second world*, we may call the world of statements in themselves the *third world*. (I now prefer to call these... 'world 1', 'world 2', and 'world 3') (UNQ 180-1).

After asking us to imagine a picture; distinguishing between the actual picture, one's mental image of it, and one's thoughts about that image; Popper used his own mental processes to illustrate the generation of a world 3 thought which, once written down, and "formulated in language so clearly *that I can look at it critically from various sides*" becomes "the thought in the objective sense, the world 3 object which I am trying to grasp.... The decisive thing seems to me that we can put objective thoughts - that is, theories - before us in such a way that we can criticize them and argue about them. To do so, we must formulate them in some more or less permanent (especially linguistic) form.... Books and journals can be regarded as typical world 3 objects..." (UNQ 182). He added, "we may include in world 3 in a more general sense all the products of the human mind, such as tools, institutions, and works of art" (UNQ 187).

Popper described world 3 somewhat paradoxically as both "man-made" and "autonomous:" "the third world, the world of objective knowledge... is man-made. But... this world exists to a large extent autonomously... it generates its own problems, especially those connected with methods of growth; and... its impact on any one of us, even on the most original of creative thinkers, vastly exceeds the impact which any of us can make upon it" (OKN 147).

### Problems

First, there seems little conjectural about the theory of worlds 1, 2 & 3. In none of Popper's several presentations is the theory offered as an hypothesis. Rather, it is laid out as a discovery, as what he thought the facts to be.

Second, the idea of objective knowledge appears directly to contradict CR. If knowledge can exist *objectively*, it is not clear how it remains at the same time *conjectural*. The exercise of studying Popper, for instance, depends on the existence of a dozen or more world 3 objects - his books. Now, either those books exist and say what they say or they don't, there is simply no room for conjecture.

Third, it not clear how we gain access to this *objective* third world when our brains and senses are 'impregnated' with inborn expectations, and are thus incapable of unadulterated contact with reality. World 3 may exist, 'out there,' objectively, but Popper said "there is no such thing as an unprejudiced observation" (OKN 51). It would therefore be difficult to know if we were actually observing world 3, or to identify what we were observing in it.

Further, when thoughts have been objectified as world 3 artifacts, it is not apparent how they accord with Popper's rejection of definitions. Once CR is part of an objective world 3, then either the words 'Critical Rationalism' correspond to the world 3 fact that there is such a scientific method, or they do not. We have a genus (scientific methods) and a species (Popper's method) whose differentia is the process of conjecture and refutation. Calling Popper's method syllogistic, or dialectical, would be manifestly wrong. Thus it would be perfectly in order, not a 'tiresome quibble,' to argue about the definition of CR with anyone who maintained, say, that conjecture and refutation was merely Logical

Positivism in disguise. Their assertion would be *untrue*; it would *not* correspond to the objective, world 3 facts.

The existence of objective world 3 ideas also seems to conflict with Popper's rejection of 'essentialism' - the real existence of concepts - which formed an integral part of his notorious attack on Aristotle<sup>29</sup> and underlay his dislike of definitions. Surely it is unreasonable on the one hand to lambast essentialism - the idea that concepts are, or have, real 'essences,' which exist in our own reality or in another dimension - while claiming on the other hand that concepts have a separate existence in world 3.<sup>30</sup>

Another awkward question must be 'Why should we stop at worlds 1, 2, and 3?' The basis for the theory is fundamental difference in kind, the worlds are "utterly different" (UNQ 181). However, in *The Open Universe*, Popper suggested the possibility of a world 4 of art (TOU 115) and a world 5 of human institutions (TOU 154). He also spoke of "the gulf which separates the human brain from the animal brain" (TOU 122). But if we are dividing reality according to fundamental differences in kind, animal consciousness ought to be world 6; and if art gets a world of its own, surely commerce is sufficiently different to qualify as world 7 - 'utterly different' things should not be left together. So plants would require a separate world from animals; elephants from amoebas; inanimate things from animate, etc. The logic of Popper's argument thus seems to lead to an Aristotelian universe of distinct entities grouped according to the identifying characteristic (or 'essence') of each kind, an inference Popper would have disliked.

Finally, the 'autonomy' of man-made, objective knowledge shows a marked kinship to Aristotle's concept of potentiality. Popper often used number theory to explain world 3: "natural numbers are the work of men," he stated. However "unexpected new problems arise as an unintended by-product of the sequence of natural numbers.... These problems are clearly *autonomous*. They are in no sense made by us; rather, they are *discovered* by us; and in this sense they exist, undiscovered, before their discovery" (OKN 160-1). That is fair enough, but is it not merely another way of saying that the future is not actual but potential; that unknown future advances do not *actually* exist, yet must exist as *potential* in the known?

In this regard it is instructive to look at Popper's idea (in physics) of "the measures of possibilities" which he called "objective probabilities" or "propensities" (TOU 105) and thought of as "*physically real*" (QTSP 133). These provide "a programme for a theory of change... which would allow us to interpret any real state of the world as both an actualisation or realisation of some of the potentialities or propensities of its preceding states, and also as a field of dispositions or propensities to realise the next state" (QTSP 198).

Leaving aside the problem of how 'physically real possibilities' fit into the category of conjectural knowledge, Popperian 'propensity' appears so similar to Aristotelian 'potentiality' - "all movement or change means the realisation (or 'actualisation') of some of the potentialities inherent in the essence of a thing" (OSE2 6) - that, in fairness, one must note that Popper dismissed Aristotle's thoughts about potentiality as "pretentious jargon" (OSE2 7).

Unfortunately, space does not permit exploration of what may be the most serious problem with the three-world theory: its well-signposted detour into

idealism, which Popper told us elsewhere he “thoroughly opposed” (OKN 323).<sup>31</sup> Suffice it to say that it is *this world* we seek to understand; and while philosophers from Plato onward have speculated about other worlds, not one of their conjectures has deepened our understanding of this one. In the words of John Searle: “We live in one world, not two or three or twenty-seven.”<sup>32</sup>

### 11. Established Theories

The last major area of difficulty with CR to be examined in this paper concerns theories which have successfully withstood criticism. Popper did allow that after scientific theories have passed a great number of severe tests, “their tentativeness may cease to be obvious” (POH 131). But if asked about ‘established’ theories he was very likely to point to Isaac Newton’s “unquestionable truths” (UNQ 37) which, seemingly unassailable for over 200 years, were pushed aside by the “Einsteinian revolution” (UNQ 81).

Yet theories do exist which, in fact, are positively confirmed, as Grover Maxwell has pointed out (PKP1 292ff). Copernicus’s heliocentric theory, for example, was indeed hypothetical in 1543 because the instruments did not then exist with which to prove it. But now that huge telescopes and space probes have eliminated any rational doubt that the earth revolves around the sun, it would seem bizarre to maintain that heliocentricity remains conjectural.

Another famous theory is that of Harvey and the circulation of the blood. Once, that was indeed a bold conjecture. But if one were to declaim nowadays that Harvey’s theory is refutable, or that we don’t know what we are talking about when we say that blood circulates in the human body, one should expect laughter from one’s audience.<sup>33</sup>

Popper was evidently aware of this problem. He once wrote about the “realisation” of the “conjecture” of an atomic bomb (TSIB 47). But if a conjecture is *realised* it is very difficult to see how it remains a conjecture. One might fairly retort, rather, that this one admission blows apart the notion of demarcation by refutability and the whole of CR along with it.

There is also the awkward subject of evolution. Popper called Darwinism “a brilliant scientific hypothesis” about “a host of biological and palaeontological observations.” He added: “I see in modern Darwinism the most successful explanation of the relevant facts” (POH 106). Later, he confirmed that he was “very ready to accept evolution as a fact” (UNQ 167).<sup>34</sup> But it is not easy to see how a ‘fact’ can be based on observations when Popper has told us that there is no such thing as an unprejudiced observation. Nor did he explain why we should suddenly accept an ‘hypothesis’ as a fact and not as a conjecture.

Popper’s problem was of course that the theory of evolution is just about as inductive as one can get, yet he wanted us to believe that induction is a myth. He found no way out of this impasse, and in the end decided that the only solution was to evade the issue: “I have come to the conclusion that Darwinism is not a testable scientific theory, but a *metaphysical research programme*” (UNQ 168).<sup>35</sup>

## 12. The Ultimate Test

CR urges us to submit our theories to severely critical tests. For a philosophy, the most critical test of all may be whether its proponents actually follow it. The example was set by Hume, who admitted that he found his scepticism hard to live by. Popper evidently experienced the same difficulty. It is easy enough to say “our scientific theories must always remain hypotheses” (OSE2 12) but much more difficult to abide by that principle consistently. Thus Popper’s use of the words ‘knowledge,’ ‘know,’ ‘truth’ and ‘fact’ often seemed to conflict with CR. He wrote, for instance: “Matter... consists of complex structures about whose constitution we know a great deal” (TOU 152-3). He urged us to pay attention to the “*invariant content or meaning*” of a theory “upon which its truth depends” (OKN 240). He referred to “universal laws” as “part of our common knowledge” (POH 145); to “objectively true” statements (TOU 119); to the ‘fact’ that “theories or expectations are built into our very sense organs” (OKN 146), and to the “undoubted” fact that “we can *learn from experience*” (C&R 291). All these assertions seem to defy, in one way or another, the idea that knowledge remains conjectural.

Popper’s philosophical premises also led him into more serious confusions. For example, he explicitly rejected as “utterly naïve and completely mistaken” what he called “the bucket theory of the mind” (OKN 61), the idea that “before we can know or say anything about the world, we must first have had perceptions - sense experiences” (OKN 341). Yet earlier he had stated: “I readily admit that only observation can give us ‘knowledge concerning facts’, and that we can... become aware of facts only by observation” (LSCD 98).

Popper’s attitude to ‘the laws of nature’ was just as perplexing. In *Open Society* he called natural law a “a strict unvarying regularity.... A law of nature is unalterable; there are no exceptions to it.... laws of nature... can be neither broken nor enforced. They are beyond human control...” (OSE1 57-58, cf OKN 196). But such absolutist claims are difficult to reconcile with the actual discovery of natural laws when, according to Popper: “There can be no valid reasoning from singular observation statements to universal laws of nature” (RASC 32, cf OKN 359).

In like vein, Popper’s use of illustrations often involved disregard of his own dicta. In *Realism and the Aim of Science*, when once again attacking induction, he told us that “mere supporting instances are as a rule too cheap... they cannot carry any weight” (RASC 130); and that, “confirming instances are not worth having” (RASC 256). However, when he had earlier sought to demonstrate the case that “practically every... ‘chance observation’ is an example of the *refutation* of some conjecture or assumption or expectation,” he unhesitatingly drew attention to scientific discoveries by Pasteur, Roentgen, Crookes, Becquerel, Poincaré and Fleming to reinforce his point (RASC 40).

The trait of employing what he sought to deny can be found throughout Popper’s work. Take his critique of Plato’s politics. In Volume 1 of *Open Society* Popper went through the *Republic*, *Laws*, etc, with a sort of remorseless philosophical laser. Yet not once did he give any hint that he regarded the object of his study as conjectural. His method was purely and simply inductive. He took Plato’s dialogues as fact, examined them line by line in search of evidence, and generalised his (very firm) conclusions.<sup>36</sup>

As an aside, it may be noted that Popper was not renowned for living up to his philosophy in his professional life. His obituary in *The Times* recorded his reputation as “a difficult man.” *The Daily Telegraph* commented, “Popper’s belief in his own infallibility was remarkable.” Later, *The Times Magazine* reported that Popper’s students at the London School of Economics found him so intolerant of criticism that they used to joke about “The Open Society by one of its enemies.”<sup>38</sup>

### Popper and Marx

Popper’s most egregious lapse as a critical rationalist concerns Karl Marx. Like so many young men of his era, Popper early embraced Marxism, but unlike so many, he also early rejected it - as an economic theory: he never discarded the Marxian ideal of social betterment for the working class, and for most of his life remained a dedicated interventionist and welfare-statist. Thus in *Open Society*, while criticising *Marxism*, he presented an almost fulsome portrait of Marx the man as a brilliantly original thinker and philanthropist, and as one of the “liberators of mankind” (OSE2 122).

In 1948, however, Leopold Schwartzschild published *The Red Prussian*. In this critical biography, based on original sources such as the Marx-Engels correspondence, Marx emerged as anything but a philanthropist. He was in fact a disgraceful sponger and drunkard, as deceitful and vindictive as he was lazy, who loathed and despised the workers (“those asses”) and whose only real animus was a deep lust for power. Nor was Marx’s thinking either original or based on original research. He borrowed most of his ideas from other socialists and his best-known thesis was pulled out of thin air without a shred of fact to support it. When he did bestir himself to try and corroborate “our view” - and found that the historical and economic data flatly contradicted him - he ignored or suppressed the evidence.<sup>39</sup>

Although Popper read *The Red Prussian* “some years” after it came out (OSE2 396), he never corrected or modified the glowing portrait of Marx he had given us in *Open Society*. It took him some 15 years even to acknowledge his awareness of the “shattering” evidence which had so drastically falsified his most famous work (OSE2 396).

In 1986, Anthony Flew, in his Introduction to a new edition of Schwartzschild’s book, gently chastised Popper for not correcting his false picture of Marx.<sup>40</sup> The publisher sent a copy to Popper, and two years later Popper wrote to Flew saying, “I wish to explain my final note (on Schwartzschild). (1) Routleges [sic] never told me in time of a new reprint. I had to squeeze things in, at the last moment. (2) I was personally shattered by Schwartzschild’s book; and it was only my view of Marx’s moral stature that was shattered. The reason that my view of Marx’s status as a scientist was not shattered is very simple: I had not had a very high opinion to start with, but I had given him all the benefit of the doubt; and my opinion had slowly deteriorated, both while writing the book and after.... it was only when I now read your Introduction that I saw I ought to have referred to my changed view of Marx’s scientific sincerity. I therefore accept your criticism fully.”<sup>41</sup>

This explanation is not really satisfactory. Popper saw the ‘shattering’ evi-

dence about Marx in the late 1940s or early 1950s, yet his “final note” was not penned until 1965. In between, there were no less than four new editions of *Open Society* in which he could have published a revised judgment of Marx. In the end, all he gave us was a reluctant, 150-word appendix on the last page of the last edition (1966).

It is also hard to accept that Popper’s opinion of Marx had not been very high. When someone writes, for example, that Marx’s theory of surplus value was “brilliant” and “a theoretical success of the first order” (OSE2 172-3); that Marx’s exploitation theory “deserves the greatest respect” (OSE2 178); and that Marx made “serious and most important contributions to social science” (OSE2 253); it does not look as though the writer’s opinion is ‘deteriorating.’

There is besides the problem that Popper later had a perfect opportunity to retract his portrait of Marx. In 1966, Professor H.B. Acton of Edinburgh University wrote that, according to Popper, “Marx was primarily concerned with achieving freedom for individual men and women” and that nothing published in the twenty years since *Open Society* had appeared required “any radical modification” of this view (PKP2 876). Yet, in his 1974 response to Acton, Popper merely pleaded guilty to having “idealized the picture of Marxism” over some minor points: there was not one word about Schwartzschild (PKP2 1162-5).<sup>42</sup>

### Conclusion<sup>43</sup>

This paper is not the first to subject Popper’s Critical Rationalism to detailed criticism. P.A. Schilpp’s *The Philosophy of Karl Popper* contains several less than sympathetic essays, as does Anthony O’Hear’s *Karl Popper: Philosophy and Problems*. And of course O’Hear earlier devoted a whole book to the matter. Other writers have been led to outright rejection. When *The Logic of Scientific Discovery* first appeared, Popper’s famous contemporary Hans Reichenbach asserted bluntly: “The results of this book appear to me completely untenable... I cannot understand how Popper could possibly believe that with respect to the problem of induction his investigations mean even the slightest advance.”<sup>44</sup>

Nonetheless, although this paper rejects Popper’s main thesis, it should not be interpreted to imply that the study of his work is valueless. Far from it. Popper wrote well and clearly, and books such as *The Open Society and its Enemies*, *The Poverty of Historicism* and *The Open Universe*, while flawed or incomplete, are full of valuable insights, astute observations, and stimulating, sometimes inspiring, prose.

A critical attitude, particularly a self-critical one, is also every bit as important in philosophy as Popper thought it was, even if he did not always exercise his own. Subjecting one’s pet theories to the kind of penetrating analysis Popper was so good at is the healthiest mental activity one can undertake. Conviction is much easier to come by than rectitude and we must always be on guard against “cocksureness” - as Popper so rightly warned us (OSE2 387).

It is also well worth keeping in mind that even if Popper was mistaken in his overall rejection of induction, CR does share with induction one of its most important elements - disconfirmation - an element which has not lost one iota of its importance since Francis Bacon first drew our attention to it in the 17th

Century. We are *not* omniscient. We *are* fallible. Disconfirming instances *must* be sought and, where not found, anticipated at any and all times.

One famous instance cited by Popper was the discovery of deuterium in water, or 'heavy' water: "Prior to this discovery, nothing more certain and more settled could be imagined in the field of chemistry than our knowledge of water... This historical incident is typical... we cannot foresee which parts of our scientific knowledge may come to grief one day" (OSE2 374-5).

There is much truth in that. But "come to grief" overstates the case. And that is where Popper went wrong: he focused on disconfirmation to the exclusion of everything else. He tried to elevate an important but isolated premise to the status of a philosophical system. Critical Rationalism is not so much a *replacement* for induction, as an exaggerated focus on the *negative element* of induction.

The Objectivist philosophy of Ayn Rand was referred to earlier. Although as unacademic as Popper was academic, Rand did share with him a number of philosophic premises; such as metaphysical realism, an opposition to idealism (including conceptual realism) and rejection of determinism and subjectivism. Indeed Wallace Matson has suggested that Rand and Popper had "much in common." His view has been partially endorsed by Robert Hollinger, who has written of "parallels" between the two thinkers.<sup>45</sup>

This paper will therefore conclude by conjecturing that when Popper said "in science there is no '*knowledge*'. . . in the sense which implies finality" (OSE2 12) what he may have been after was Rand's insight that concepts are *open-ended*.<sup>46</sup>

For if *Rand* had been confronted with Einstein's rewrite of Newton; or a black swan where there had only been white ones; or the discovery of a new kind of water; she would not have said, as Popper did, that our previous knowledge had been "overthrown" or had "come to grief" or that "the belief in scientific certainty... is just wishful thinking" (OSE2 374). Rather, she would have said simply that our knowledge had been *expanded*.

The description of concepts as 'open-ended' does appear to be the Philosopher's Stone which Popper sought but never found. He correctly saw that there was a problem with most people's idea of certainty, yet never quite fought his way through to an acceptable solution.

But be that as it may. Whatever one may think of Popper, or of Rand, the open-endedness of concepts certainly seems to be a more fruitful, less fraught, and more commonsensical qualification of certainty than, "We never know what we are talking about."

### Sources and Acknowledgements

The works consulted for this paper were: *The Open Society and its Enemies*, Volumes 1 and 2 (OSE1/2);<sup>47</sup> *The Poverty of Historicism (POH)*;<sup>48</sup> *The Logic of Scientific Discovery (LSCD)*;<sup>49</sup> *Conjectures and Refutations (C&R)*;<sup>50</sup> *Objective Knowledge (OKN)*;<sup>51</sup> *The Philosophy of Karl Popper*, Books 1 and 2 (PKP1/2);<sup>52</sup> *Unended Quest (UNQ)*;<sup>53</sup> *Realism and the Aim of Science (RASC)*;<sup>54</sup> *The Open Universe (TOU)*;<sup>55</sup> *Quantum Theory and the Schism in Physics (QTSP)*;<sup>56</sup> *The Self and Its Brain (TSIB)*;<sup>57</sup> and *A World of Propensities (AWP)*.<sup>58</sup>

To reduce the number of endnotes most references have been given in square brackets in the text, in the form of initial letters from Popper's titles followed by a page number. Thus the quotations in the introduction, annotated (PKP2 977) and (OSE2 39), are from *The Philosophy of Karl Popper*, Book 2, page 977, and *The Open Society and its Enemies*, Volume 2, page 39. Re style, double inverted commas are used for actual quotations, single ones for emphasis. Where a quotation begins a sentence, initial letters are sometimes capitalized to assist readability. Italics are in the original.

Lastly, the author would like to express his sincere thanks to David Conway, Anthony Flew, David Kelley, Tibor Machan and David Miller for valuable observations or criticisms which led to the reworking of many passages; to Kevin McFarlane and Brian Micklethwait for encouragement and practical help; and to The Estate of Karl Popper for kind permission to reproduce copy-right material.

#### NOTES:

1. This paper is an abbreviated and amended version of Nicholas Dykes, *A Tangled Web of Guesses: A Critical Assessment of the Philosophy of Karl Popper* (London: Libertarian Alliance, 1996).
2. In obituaries in *The Guardian* and *The Independent*, 19 September 1994.
3. E. Freeman and H. Skolimowski noted that Peirce anticipated some of Popper's central ideas (PKP1 464ff). Popper acknowledged this (PKP2 1072) though he did not read Peirce until the 1950s. He added "I feel proud of so eminent a predecessor" (PKP2 1119). The author is indebted to David Conway for pointing out that Peirce coined the word 'fallibilism.'
4. Bartley developed 'Comprehensive Critical Rationalism.' See *Evolutionary Epistemology, Theory of Rationality, and the Sociology of Knowledge*, G. Radnitsky and W.W. Bartley III Eds (Lasalle, Illinois: Open Court, 1987). Cf I.C. Jarvie's review, *Critical Review* 2/1.
5. David Miller, *Critical Rationalism: A Restatement and Defense* (Lasalle, Ill: Open Court, 1994). Though vigorously written, this book did not seem to solve the problems raised by this paper, nor others raised by Anthony O'Hear in *Karl Popper* (London: Routledge and Kegan Paul, 1980). See also Anthony O'Hear, Ed, Karl Popper: *Philosophy and Problems* (Cambridge, UK: Cambridge University Press, 1996).
6. The proposition is used as a first premise by Jan C. Lester in *Escape from Leviathan: Liberty, Welfare, & Anarchy Reconciled* (London: Macmillan, forthcoming 1999). Well-known scientists who acknowledged Popper's influence include Jacob Bronowski, John Eccles, Peter Medawar and John Maynard Smith.
7. David Hume, *An Enquiry Concerning Human Understanding*, 3rd Edn, L.A. Selby-Bigge & P.H. Nidditch Eds (Oxford: Oxford U.P. 1975), p. 74.
8. *Ibid.*, p. 96.
9. *Ibid.*
10. David Hume, *A Treatise of Human Nature*, Book 1 (London: Fontana/Collins, 1975), p. 63.
11. H.W.B. Joseph, *An Introduction to Logic*, 2nd Edn (Oxford: OUP, 1916), p. 408.
12. *Ibid.*, p. 408.
13. *Ibid.*
14. Popper's notion of innate cerebral content is deeply problematic but space permits no discussion. For details, see OKN 26-7, 63, 71-2, 258; C&R 27, 47-8; TSIB 116, & AWP 37, 46.
15. Brand Blanshard, *Reason and Analysis* (Lasalle, Ill: Open Court, 1964) p. 39. See also Nathaniel Branden, "The Contradiction of Determinism," *The Objectivist Newsletter*, May 1963.
16. Anthony Flew, *Hume's Philosophy of Belief* (1961, 1966; Bristol, UK: Thoemmes, 1997), p. 31. The author is indebted to Professor Flew for this point.
17. J.W.N. Watkins noted that if our senses were actually unreliable we wouldn't be here (PKP1 404). Popper half acknowledged this (PKP2 1114, AWP 32).
18. 'Knowledge remains conjectural' resembles the traditional sceptic's claim that "knowledge is impossible," which is an obvious self-contradiction. Popper allowed that "the term 'conjectural knowledge' may be claimed to be a contradiction in terms" (OKN 76), but did not explain why he chose to use it. Imre Lakatos observed: "The difference between total scepticism and humble fallibilism is so small that one frequently feels that one is engaged in a mere verbal quibble" (PKP1 260).
19. Nathaniel Branden, "The Stolen Concept," *The Objectivist Newsletter*, January, 1963.
20. Popper touched on this: "the growth of knowledge consists in the modification of previous knowledge" (OKN 71). He attempted to resolve the infinite regress by positing 'inborn dispositions and expectations.' See note 14.

21. Tom Settle has made the same complaint. On the question of which hypothesis to choose he wrote, "we get no good guidance from Popper" (PKP2 702).
22. Einstein was Popper's "hero" (AWP 8). There are over 120 references to him in just three books: LSCD, C&R and OKN.
23. Cf Anthony O'Hear, *Karl Popper*, op cit, p. 97.
24. George H. Smith, *Atheism: the Case against God* (Buffalo NY: Prometheus Books, 1989), p. 51ff.
25. *Karl Popper: Philosophy and Problems*, op cit, p. 4.
26. Brand Blanshard noted that particular propositions such as 'some swans are white' can only be falsified by showing that 'no swans are white;' *Reason and Analysis*, op cit, p. 228. This creates a serious difficulty for CR, but the problem has been ignored by Popperians such as Bryan Magee, whose *Popper* (London: Fontana, 1973, 1985) has maintained through ten editions that: "Popper's seminal achievement has been to offer an acceptable solution to the problem of induction" (p. 22).
27. The same point has been made by Peter Lipton: "There is no reliable route to falsification that does not use induction..." "Popper and Reliabilism," *Karl Popper: Philosophy and Problems*, op cit, p. 43. The author is indebted to Dr Lipton for sharpening his focus on this issue.
28. Cited by John Herman Randall, Jr., *Aristotle* (New York: Columbia University Press, 1960), p. 116.
29. OSE2 1-26. See *Tangled Web*, op cit, pp. 18-20, for a detailed examination of Popper's odd critique.
30. Popper saw this weakness and later referred to worlds 12&3 as "modified essentialism" (PKP2 1115). For further criticism see O'Hear, *Karl Popper*, Ch. IX.
31. See *Tangled Web*, op cit, pp. 22-4, for a discussion of Popper's 'idealism.'
32. John Searle, *The Mystery of Consciousness* (London: Granta Books, 1997), p. 88.
33. Cf C&R 41n8: "most dissectors of the heart before Harvey observed the wrong things - those which they expected to see." But this is to argue for Baconian objectivity, *against* CR. Harvey broke free in a way Popper thought impossible.
34. Confusingly, Popper also wrote that Darwinism, "has very little content and very little explanatory power, and it is therefore far from satisfactory... we should try hard to improve upon Darwinism, or to find some alternative" (PKP2 1084). He may have meant Lamarck; cf RASC 94 & TSIB 425.
35. Cf John Maynard Smith: "[It is] an occupational risk of biologists to claim, towards the end of their careers, that problems which they have not solved are insoluble." *Did Darwin Get it Right? Essays on Games, Sex, and Evolution* (London: Penguin, 1993), p. 249. See also UNQ 187: "theories which deny what they cannot explain," and TOU 151, "The solution... is the denial." Popper and induction?
36. For criticism of Popper's view of Plato see R. Levinson, *In Defense of Plato* (Cambridge Mass: Harvard U. Press, 1953); and J. Wild, *Plato's Modern Enemies and the Theory of Natural Law* (Chicago: U. of Chicago Press, 1953), also PKP2 859. For Popper's replies, see OSE1 323 and PKP2 1159.
37. Both obituaries appeared on 19 September 1994.
38. Jim McCue, "Mind Reading," 13 May 1995, p. 21. Confirmed by Frederick Raphael in *Popper* (London: Phoenix, 1998), p. 12.
39. Antony Flew, "Introduction to the 1986 Edition," in Leopold Schwartzschild, *The Red Prussian* (London: Pickwick Books, 1986), p. 7.
40. *Ibid*, pp. 3-4.
41. Karl Popper to Anthony Flew, private letter, 1988. The author is deeply indebted to Professor Flew for providing a copy.
42. For a detailed discussion of Popper's attitude to Marx, see *Tangled Web*, op cit, pp. 26-29.
43. Since this paper rejects CR on logical grounds, it has not been thought necessary to discuss buttress concepts such as 'degrees of testability,' 'corroboration,' 'verisimilitude,' etc. These are analysed in O'Hear, *Karl Popper*, Ch. 3, and defended by David Miller, *Critical Rationalism*, Ch. 10 and passim. For accounts of Popper's ethics and politics, see *Tangled Web*, op cit, pp. 24-26 & 29-30.
44. Quoted on the dust jacket of the 1972 Hutchinson edition.
45. Wallace Matson, "Rand on Concepts," and R. Hollinger, "Ayn Rand's Epistemology in Historical Perspective," in *The Philosophic Thought of Ayn Rand*, D. Den Uyl & D. Rasmussen Eds. (Urbana, Ill: University of Illinois Press, 1984), pp. 22 & 56.
46. "... a concept is an 'open-end' classification, which includes the yet-to-be-discovered characteristics of a given group of existents." Ayn Rand, *Introduction to Objectivist Epistemology* (New York: The Objectivist Inc, 1967), p. 60.
47. London: Routledge and Kegan Paul, 5th Edn, 1966.
48. London: Routledge and Kegan Paul, 1961.
49. London: Hutchinson, 1972.
50. London: Routledge, 1989.
51. London: Oxford, 1972.
52. P.A. Schilpp Ed, *The Library of Living Philosophers*, Volume XIV, Books 1 & II (Lasalle, Illinois: Open Court, 1974).
53. London: Fontana/Collins, 1976.
54. London: Routledge, 1992.
55. London: Hutchinson, 1982.
56. London: Hutchinson, 1983.
57. ohn C. Eccles, Karl R. Popper, *The Self and Its Brain: An Argument for Interactionism* (London: Routledge and Kegan Paul, 1983).
58. Bristol, UK: Thoemmes, 1990.

# Is Virtue Only a Means to Happiness? An Analysis of Virtue and Happiness in Ayn Rand's Writings<sup>1</sup>

Neera K. Badhwar

## Introduction<sup>1</sup>

"Virtue is not an end in itself. Virtue is not its own reward . . . .  
*Life* is the reward of virtue—and happiness is the goal and  
reward of life."<sup>2</sup>

This formulation suggests that happiness is something entirely external to virtue, a further consequence of acting virtuously. Virtue, on this view, is only an instrumental means to the agent's happiness. As Leonard Peikoff states in his book on Objectivism, "[i]n the Objectivist approach, virtue is (by definition) the means to value," including the supreme value, happiness. Virtue is practical, he explains, in the sense that it "minimizes the risks inherent in life and maximizes the chance of success" or happiness (328).

However, although this is Ayn Rand's official view, she does not always treat virtue as purely instrumental to happiness. As I will show, her novels and some of her theoretical statements present a different view, a view that, I believe, is far closer to the truth. Unfortunately, the purely instrumental analysis of virtue has become standard in current interpretations of Objectivist ethics, thanks to the persistence of two false assumptions. One assumption is that the sole alternative to regarding virtue as merely instrumental to happiness is to regard it as wholly an end in itself, i.e., as Rand puts it, as "its own reward". Another is that to regard virtue as an end in itself is to regard it as quite unconnected to happiness. And this is to open the flood-gates to the irrationalism of intrinsicism or supernaturalism. Hence, the consequence of rejecting virtue as merely instrumental to happiness is to be unable to justify virtue in rational terms.

However, both assumptions are false. First, the alternative to regarding virtue as merely instrumental to a further end is not necessarily to regard it as wholly an end in itself. There is a third logical possibility, namely, to regard virtue as partly a means to happiness and partly an end in itself. Further, to regard virtue as an end in itself is not necessarily to regard it as unconnected to happiness. This is, indeed, how Kant regarded it, but not, for example, Socrates or the Stoics. It can be an end in itself in the sense that it is (wholly or partly) *constitutive* of the supreme end, happiness. I believe that conceiving of virtue as purely instrumental to happiness shows a misunderstanding not only of the nature of virtue, but also of the nature of happiness. An adequate analysis of the virtues requires that we recognize virtuous activity as an ineliminable

constituent of happiness, and an adequate analysis of happiness requires that we recognize it as partly constituted by virtuous activity. This conception of the relationship between virtue and happiness allows happiness to remain the *summum bonum*, while leaving room for justifying virtue in terms of its role in happiness. This conception of virtue and happiness is also the one that best captures the vision of the ideal individual—the individual of virtue—and of the ideal life—the life of happiness<sup>4</sup>—in Rand’s novels. And it is implied by at least some of her explicit statements about the relationship between virtue and happiness. In this paper I will give an analysis of virtue, of happiness, and of their relationship that is both philosophically defensible, and adequate to Rand’s vision of the ideal individual and the ideal life.

## THE NATURE OF VIRTUE

### Rand’s Conception of Virtue

Rand defines virtue as the act by which we gain/and or keep value.<sup>5</sup> But she also defines particular virtues, such as justice, pride, integrity, honesty, et al, more fully in terms of the recognition of certain facts and of actions that accord with such recognition. Thus, justice “is the recognition of the fact that you cannot fake the character of men as you cannot fake the character of nature, that you must judge all men as conscientiously as you judge inanimate objects, with the same respect for truth, with the same incorruptible vision, by as pure and as *rational* a process of identification—that every man must be judged for what he is and treated accordingly . . .” (AS, 937, FNI, 129). Similarly, integrity “is the recognition of the fact that you cannot fake your consciousness” (AS, 936, FNI, 129), a recognition that is expressed in loyalty to one’s rational values and convictions in the face of the contrary opinions of others (VOS, 28, 52, 80). And honesty “is recognition of the fact that you cannot fake existence,” a recognition that is expressed in truthfulness in thought and speech (AS, 936-37, FNI, 129).

When Rand says, “you cannot fake the character of men”—or your consciousness or existence—she obviously does not mean that it is impossible to do so, since this would imply that injustice or lack of integrity or honesty are impossible. She means that you cannot do so in the long run without detriment to yourself, that to do so is disvaluable. Thus, recognition of the value of not faking various aspects of reality in thought and deed—or, in positive terms, of facing reality—is implicit in virtuous action. When we act virtuously, whatever other values we might aim to bring about, we give expression to—and, thereby, maintain—the value we place on facing reality. In this sense, every virtuous action both maintains a value, and is a means to some value. This is in keeping with Rand’s general definition of virtue as the act by which we gain or keep value.<sup>6</sup>

The value of justice, integrity, and honesty, as of the “higher-order” virtues of rationality, productiveness, and pride, is connected to what Rand regards as the three cardinal values: reason, purpose, and self-esteem. These values, says Rand, are “the means to and the realization of one’s ultimate value, one’s own life” (VOS, 27) as a rational being, and, therefore, one’s own happiness, conceived of as a “successful state of life” and its emotional concomitant (AS, 932; FNI, 123; VOS, 27-29).<sup>7</sup>

There is, thus, a hierarchy of values, as there is a hierarchy of virtues. There are the specific values connected to the different virtues, the three cardinal values, and the ultimate value, happiness. The particular values have a necessary connection to the three cardinal values, and these to happiness. As Rand says, the cardinal values are both the means to, and the realization of, one's ultimate value, happiness. As far as I know, Rand does not explain what it means for these values to "realize" happiness, or how they do so. But when we talk of an action or state of affairs realizing something, we mean that it gives expression to, or embodies, that thing. Thus, a process of self-realization is a process of giving expression to the self, of "bringing forth" the deepest aspects of one's self. Again, a career that realizes one's aspirations is a career in which one can give expression to one's aspirations, embody them in one's work. Applying this to the cardinal values, then, we can say that reason, purpose, and self-esteem realize an aspect of happiness because they express or embody an aspect of happiness. Thus, self-esteem is a sense of self-worth, one's worth as a person, and the state of having self-esteem is inherently—by its very nature—a state of deep, enduring satisfaction. But this is exactly the sort of state we think a happy life must include. It follows, then, that self-esteem is itself partly constitutive of happiness. It is also a means to happiness because the sense of one's worth as a person serves as an important motivating factor in acting to achieve happiness.

Putting Rand's definitions of virtue together, we can say that, according to Rand, virtue consists of recognizing various values as both means to, and part of, happiness, and acting to gain and/or keep them.

Even this fuller definition, however, will not quite do. What is missing is the idea that a virtue is a character trait, an enduring disposition or orientation that is expressed in virtuous acts. As Rand's novels amply illustrate, our moral responses reveal our characters—our selves, our souls. And our characters consist not only of particular cognitions of value and actions motivated by such cognition, but also of general dispositions or tendencies to so cognize and act.

But even this is not enough. Rand's language often suggests that the recognition of values that is part and parcel of virtue is entirely intellectual in nature. But virtuous character traits are not only intellectual dispositions to apprehend and achieve value, they are also emotional dispositions. The rationality of virtuous dispositions and actions, I will argue, is a function of the intellect as well as of the emotions. Hence, when I refer to a virtuous disposition as a rational disposition, I will have in mind an integrated intellectual and emotional disposition. It is this sort of disposition that is possessed by Rand's protagonists, whom she sees as exemplars of virtue, of moral excellence. In the next section, I will outline a conception of virtue that captures the character of Rand's protagonists better than her own explicit statements about virtue, and that is more adequate to our everyday and scientific knowledge of human psychology.

### **A More Adequate Conception of Virtue**

What must be true of virtuous traits and actions if they are to count as morally excellent, the pinnacle of moral achievement?

(i) First, to count as excellent, a virtuous act must not only be motivated by a

particular cognition and choice of the truly valuable, it must also express a standing disposition or habitual tendency to cognize and choose what is truly valuable. For an act that expresses a standing disposition is more deeply rooted—and, thereby, better—than an act that is merely motivated by a particular cognition.

(ii) Secondly, to count as excellent, virtuous traits must make us responsive to the morally relevant features of the situations we face. But someone whose emotional *dispositions* are at variance with her intellectual dispositions will often fail to notice the morally relevant or important features of a situation. And so she will be a less reliable moral agent than someone whose emotions are integrated with her intellectual convictions.

The idea that (irrational) emotions can disrupt rational thought and action is a commonplace. But the idea that (rational) emotions are required for rational thought and action is simply the other side of the same coin. Depending on whether one's emotions are rational or irrational, they will direct one's attention towards or away from what is truly important and, thereby, affect the accuracy of one's total picture of things. Hence, someone who is committed to doing the right thing and has the right principles, but whose emotions are at variance with her intellectual commitments, will often fail to notice exactly what sort of response justice or courage or kindness requires in a particular instance. For intellectual principles alone cannot tell us what is relevant or important to one's choice of action in every particular situation of a certain kind, any more than medical principles can tell us which symptoms are relevant or important to the right diagnosis in every instance of a certain disease. The morally important features of a situation depend on the current and past context, and contexts vary indefinitely. For example, a principle, or set of principles, can tell us that when someone has suffered a loss through his own carelessness, sometimes the important or relevant feature of the situation is the loss (and the right response sympathy), at other times, the carelessness (and the right response probably something other than sympathy). But principles can provide only this sort of general guidance; they cannot tell us which feature is relevant or important when.<sup>10</sup> The ability to discern what is relevant or important in a given situation depends, in part, on experience and the stock of value-judgments that are embodied in our (rational) emotions.

A vast amount of both everyday and scientific evidence supports this point. It also supports the more general and basic point that it is emotions that make us aware of the value-dimension of most things in the first place and, indeed, partly constitute many of our values. If human beings lost their emotional faculties and became beings of pure intellect, they would also lose most of their values or their ability to apprehend values. Thus, because he is largely intellect, Star Trek's Spock can neither see the importance of certain things in human life, nor have many of the same values. To paraphrase Daniel Goleman, without emotions the intellect is blind (*ibid*, 53). Likewise, people with an impaired emotional faculty, such as psychopaths, or people who have suffered certain sorts of brain injuries, are unable to grasp what matters in human affairs. They are rational in a purely abstract sense: they can perform complex calculations and deductions, and can even follow arguments for doing or not

doing certain things. But they simply cannot be motivated by their abstract intellectual understanding of what must be done to attain certain ends, because these ends mean nothing to them, have no *importance* to them. In standard philosophical terminology, they have *theoretical* rationality, but no *practical* rationality, neither in connection with their own welfare, nor in connection with others' welfare. In the case of psychopaths, at least, this inability to have a sense of the importance of things leads to a profound amorality.

The idea that a virtuous disposition must include not only an intellectual commitment to objective values, but also an emotional orientation towards such values, is well illustrated in Rand's depiction of her characters. The following passage from *Atlas Shrugged* shows how the emotions of someone who possesses the virtues enhance her awareness and guide her responses.

Dagny, the heroine of the novel, has been looking for a scientist who can understand the design and structure of the motor she has discovered in a scrap pile, the motor she later learns was invented by Galt. On failing to find anyone intelligent enough or interested enough in her discovery, she reluctantly calls upon the brilliant Dr. Stadler. Reluctantly, because, despite his dedication to principles of rationality and truth in science, he fails to apply them to human affairs. As he has told Dagny on an earlier occasion, "[m]en are not open to truth or reason," and must be deceived or forced if the men of intellect are to accomplish anything (180). And so he endorses the establishment of a state-funded Institute of Science, and allows himself to become a lackey of politicians in the name of saving science. When Hank Rearden's metal is unjustly attacked in his name, he refuses to dissociate himself from the attack. This is the background of Dagny's decision to meet with Dr. Stadler in the hope of uncovering the secret of the motor—and its inventor.

When Stadler reads about the motor in the materials that Dagny presents to him, he openly expresses his astonishment and delight at the extraordinary achievement. Dagny wishes that "she could smile in answer and grant him the comradeship of a joy celebrated together," but finds herself unable to do any more than nod and say a cold "Yes" (332). Her response here is true to the full context of her knowledge of Stadler, a context made immediately available to her only with the help of her emotions. Throughout the discussion her responses are guided by her knowledge of Stadler's past, even as they are finely calibrated to variations in Stadler's present behavior. Thus, when he exclaims, "It's so wonderful to see a great, new, crucial idea which is not mine," and asks her if she has ever felt a "longing" for someone she "could admire," she softens and tells him that she's felt it all her life (335).

Not only do the emotions of someone who possesses the virtues guide moral perception and response, they even sometimes correct our intellectual judgments. Thus, when Dagny is on her way to confront Francisco who, apparently, has turned into a playboy, destroying people and fortunes, she is determined to grant him no personal response, for she is certain that he deserves none. Yet when he smiles at her, "the unchanged, insolent, brilliant smile of his childhood," and greets her with their childhood greeting, she finds herself greeting him likewise, "irresistibly, helplessly, happily" (114). Her emotions pick up something that her intellect alone could not, and lead her to respond appropriately to the facts, though contrary to her intentions.

These and similar passages illustrate some of the ways in which Rand's portrayal of virtue in her novels goes beyond her theoretical statements about virtue.

To summarize the discussion thus far: to count as a moral excellence, a fully virtuous act must be deeply rooted in us, i.e., in a virtuous character trait, and such a trait must be an integrated intellectual-emotional disposition that enables us to recognize, and respond appropriately to, the relevant features of a particular situation. What else must be the case for virtuous acts and traits to count as virtuous—as the pinnacle of moral achievement?

(iii) Thirdly, a virtuous act is an act that is done not only for the right reasons—i.e., for the sake of the good, the valuable—but also in the right manner. This, too, implies that a virtue is an integrated intellectual-emotional disposition. For if our emotions are at variance with our intellectual dispositions and judgments, then, even if we recognize that a certain sort of act is called for, and why, we may fail to do it in the right manner.

For example, conceding a point in an argument when we recognize that it is only fair to do so does not count for much if we concede it in a resentful manner, and is not necessarily better than not conceding it at all (“O.k., o.k., you win!”). Again, helping someone in need when we judge that we should is not an act of kindness if we do it with an air of performing a painful duty. Nor is it necessarily better than not helping at all. Thus, the wrong manner can undermine the very rightness of an act done for the right reasons, and the manner can be wrong *even when the agent recognizes the importance of acting in the right manner*. For wayward emotional dispositions, emotional dispositions that are contrary to one's rational intellectual beliefs and commitments, can subvert one's intended responses. But even if someone with such dispositions always manages to act in the right manner through sheer strength of will when she can see what the right manner requires, she will sometimes be unable to see what it requires. And so, even though admirably strong, she will remain a less reliable—and so less good—agent than a virtuous person.

In short, fully virtuous acts express deep-seated dispositions to think, feel, desire, and respond fittingly, with fine discrimination, in a variety of situations. Since these dispositions involve the agent's emotions as well as intellect, virtuous acts express not only the agent's *commitment* to the right, but her wholehearted *love* of the right. This wholeheartedness is exemplified in Dagny's character, whose “love of rectitude,” we are told, was “the only love to which all the years of her life had been given” (AS, 512). When the “moratorium on brains” is announced, this love expresses itself in a total, cold anger—and a calm, full, intellectual certainty in the decision that she must immediately resign from the Vice-Presidency of Taggart Transcontinental (*ibid*). Only a wholehearted love of the good—a love in which all of the agent's self is involved, rather than only her intellectual self—can express virtue, because a wholehearted love of the good is better than a half-hearted or divided love. And this not only because it is more reliable, but also because it is more expressive of the worth of its object.

(iv) Lastly, as a moral excellence, a virtuous character must put us in the best state for achieving the supreme value, happiness, conceived of as a “successful state of life” and its emotional concomitant. To do this it must (a) enable us

to stay in touch with reality, and (b) integrate and harmonize our inner life. The more “gappy” our grasp of reality, the more precarious our happiness, and inner conflict is both inherently unpleasant and an obstacle to this grasp. The connection of virtue with happiness is one more reason why virtue must be seen as an integrated intellectual-emotional disposition. For, as we have already seen, inner harmony and a solid connection with reality both require an integration of our emotions with our reason.

In short, if moral virtue is excellence of character, then a virtuous disposition must be one that incorporates both our intellectual and our emotional attitudes. This is explicitly recognized by Aristotle in his definition of virtue, a definition that captures what Rand depicts in her fiction far better than her own definition.

### **Aristotle’s Conception of Virtue and Rand’s Virtuous Individuals**

Aristotle defines virtue as “a state [of character] concerned with choice, lying in a mean relative to us, this being determined by reason and in the way in which the man of practical wisdom would determine it.”

Virtue is a disposition to choose the “mean” in the sense that it is a disposition to choose the “intermediate” or appropriate response, and to do so in a wide variety of situations. By contrast, vice is a disposition to choose the “extreme” or inappropriate response. For example, the virtue of generosity is the mean opposed to the vices of prodigality and stinginess. Likewise, courage is the mean opposed to the vices of recklessness and cowardice. Someone who has the virtues has the ready ability to “hit the nail on the head”—to respond exactly appropriately—in a wide variety of difficult situations. And the disposition to respond appropriately is the disposition to feel, deliberate, choose, and act “at the right [appropriate] times, about the right things, towards the right people, for the right end, and in the right way” (NE, 1106b21-23).

Further, the mean is “relative to us” in the sense that the right or appropriate action depends on both the external circumstances of action and on certain features of the agent. Thus, what counts as generosity for a graduate student might be stinginess for a millionaire and prodigality for an undergraduate student. For example: a \$50 contribution to an organization that promotes the cause of freedom might be generous for a graduate student, prodigal for an undergraduate student, and downright stingy for a millionaire who professes dedication to the cause of freedom above all other causes. The mean or virtuous act in a given situation is “determined by reason” in the sense that practical reason—reason as applied to the question of how to act or, more generally, how to live—takes all the relevant facts into consideration.

The man of practical wisdom exemplifies practical reason at its best. For practical wisdom—the virtue of practical reason—just is excellence in practical reasoning. But practical reason both shapes, and is shaped by, emotion. Hence, practical wisdom is possible only with the proper emotional dispositions that are part and parcel of virtue. A wise and virtuous choice, Aristotle remarks, expresses “truth agreeing with correct desire” (NE 1139a30) or correct desire combined with correct thought (NE 1139b5). Thus, practical wisdom and virtue imply each other. The inner states and actions of the virtuous or wise man display not merely an intellectual commitment to principle, but an

intellectual and emotional disposition that informs his characteristic ways of deliberating, perceiving, feeling, desiring, and acting. Moreover, since we all get pleasure from doing what we love—the philosopher from philosophizing, the painter from painting, the runner from running—the person who loves virtue gets pleasure from acting virtuously. This sort of pleasure is inherent in virtuous activity, and inseparable from it. Hence, the pleasure of virtuous activity qua virtuous, including the pleasurable awareness of oneself in such activity, is not interchangeable with other sorts of pleasure, such as the pleasure inherent in running qua running, or solving a puzzle, or consuming fine truffles. These pleasures are independent of whether or not the activities in which they inhere are compatible with one's overall happiness. The pleasure of virtuous activity, by contrast, is the distinctive pleasure of tracking and expressing particular values in an awareness of their relationship to the supreme value, one's own happiness.

This does not mean that there can be no pain attendant on virtuous action. When a serious loss of, or serious damage to, other goods is involved, Aristotle recognizes that the right action will involve pain. But the pain will be due to the loss of real, important goods, not to the loss of trifles, or of things that should never have been valued in the first place. Nor, of course, will the pain come from the knowledge that one is doing the right thing—only the very vicious would find *this* painful.

Aristotle distinguishes between the virtuous man and the strong-willed or continent (*enocratic*) man. Both have the right principles and commitments, and dispositionally act in accordance with their intellectual judgment. Nevertheless, the strong-willed man falls short of practical wisdom and virtue because his emotions conflict with his intellectual judgment. He is rational without possessing that excellence of practical reason which is practical wisdom, and he is rightly motivated without possessing that excellence of desire and feeling which is virtue of character. Hence, he also lacks the fine-tuned perceptiveness and responsiveness that is characteristic of the virtuous. And he is robbed of the pleasure that the man of virtue gets from acting virtuously. He would be a better, as well as a happier, man if he were virtuous rather than merely strong-willed.

Rand does not make the Aristotelian distinction between a virtuous and a merely *enocratic* character in her ethical theory, nor are there any *enocratic* characters in her novels. But her portrayals of her ideal characters illustrate the Aristotelian conception of a virtuous character. When her heroes and heroines act honestly or fairly or kindly, they do so wholeheartedly, i.e., without inner conflict over whether to do the right thing or take the easy way out. Their choices and actions express their intellectual as well as emotional states. They *desire* to do what they correctly *perceive* as good and *intellectually believe* they ought to do. And so their responses “hit the mean” in a wide variety of situations.

A good example of this occurs in a scene in *The Fountainhead*, where Peter Keating goes to see Howard Roark to bribe him for remaining silent about his contribution to the Cosmo-Slotnick building, the building for which Keating has won an award. In the conversation that precedes the actual offering of the bribe, Keating tries to persuade Roark to compromise his principles and aim for success. “Just drop that fool delusion that you're better than everybody else—and go to work . . . Just think, Howard, think of it! You'll be rich, you'll

be famous, you'll be respected, you'll be praised, you'll be admired—you'll be one of us!"<sup>13</sup> Roark looks at him, with eyes that are "attentive and wondering," knowing that Peter is sincere, but also that he is disturbed by something in him, Roark, and asks, "Peter, what is it that disturbs you about me as I am?" (192). Keating responds honestly, acknowledging that he is disturbed by something in Roark, although he doesn't know what. In the face of this confession, Roark's response "hits the mean" by being exactly appropriate to the situation. "Pull yourself together, Peter," said Roark gently, as to a comrade. "We'll never speak of that again." To the extent that Keating is honest, he is Roark's equal, to be treated with respect, not scorn. And because he is honest and willing to show that he is ashamed of himself, he deserves the kindness of being given the chance to "pull himself together," to recover his dignity. In the next moment, however, Keating's attitude changes. He pretends that he was "only talking good plain horse sense," thereby implicitly denying his fear of Roark. Roark's attitude changes immediately: he responds to this dishonesty harshly, telling Peter to shut up. Once again, Roark's response "hits the mean," giving Peter exactly the treatment he deserves.

In this scene, as in many others, we see an individual whose responses are appropriate to the situation in all the ways delineated by Aristotle: in aim, in timing, in the emotions felt, and in manner. Such "fine-tuning" of his responses is possible only because they are informed by both his intellect and his emotions.

Rand also depicts the pleasure, or at least the sense of inner satisfaction and fulfillment, that a virtuous person gets from doing the right thing—without forgetting the painful, even tragic, aspects that the choice of the right action can involve. In *Atlas Shrugged*, Francisco's choice to give up Dagny and his work, the things he loves most, perhaps forever, for the sake of joining the strike, is a particularly dramatic example of the agonizing loss that the choice to do the right thing can involve. It is also an example of the serenity and fulfillment attendant on such a choice. On his last night with Dagny, at the height of his despair, Francisco turns to her and begs her to help him refuse Galt's call, "[e]ven though he's right" (AS, 111). By the next morning, however, after he has emerged from his agonized struggle and made his decision, his face shows "both serenity and suffering," and he looks like a man "who sees that which makes the torture worth bearing" (AS, 112).

The veridicality of Rand's portrayal of her ideal characters lends support to Aristotle's conception of virtue, just as the independent plausibility of Aristotle's conception of virtue lends support to Rand's portrayal of her characters. Aristotle's conception of vice—the worst possible state of character—is also illustrated in Rand's fiction. According to Aristotle, vice disposes an individual to feel, deliberate, choose, and act wrongly. Vice blinds a person to the good, and may even reverse his perception of good and bad, so that he sees the good as bad and the bad as good (NE, Bk. III, ch. 4). Vice, says Aristotle in a memorable phrase, is unconscious of itself (NE, 1150b 35).

This conception of vice captures Rand's portrayal of her wholly or partly vicious characters. In *The Fountainhead*, Gail Wynand is time and again shown revealing his lust for power over others without any awareness that what he is revealing is a vice—more precisely, without any awareness that lust for power over others is a vice even if, as he claims, these others are devoid of integrity.

Power, Dominique. The only thing I ever wanted. To know that there's not a man living whom I can't force to do—anything. Anything I choose . . . . They say I have no sense of honor, I've missed something in life. Well, I haven't missed very much, have I? The thing I've missed—it doesn't exist (497).

Rather like a latter-day Thrasy-machus, the anti-moralist in Plato's *Republic*, who sees the ability to be unjust when one can get away with it as a sign of superior strength, Wynand sees his ability to break people's wills as a sign of his self-sufficiency and superiority. And again rather like Thrasy-machus, who "unmasks" justice as simply a ploy of the strong to get the weak to serve their interests, Wynand "unmasks" people's belief in integrity as simply another expression of their dishonesty, interpreting his own cynicism as a sign of his clear-sightedness and honesty (497).

However, contrary to Aristotle's suggestion, the vicious are not *always* unconscious or ignorant of their vice. Sometimes they are aware of their vices as vices, but, as Rand emphasizes, they evade this knowledge, as they evade knowledge of many other facts. Sometimes, again, habitual evasiveness combines with ignorance to put a person at the mercy of his vicious dispositions, which then "break through" and subvert his better intentions, even to his own detriment. Consider again the scene where Keating goes to see Roark to bribe him for remaining silent about his contribution to the Cosmo-Slotnick building. He has "planned the interview to be smooth and friendly," with a manner to match (FH, 191). But he surprises himself by starting off with the words, "What's the matter, Howard? You look like hell. Surely, you're not overworking yourself, from what I hear?" (191). His manner is insultingly familiar and condescending, prompted by his desire to show Roark that he is not afraid of him—a desire that overcomes the intention to conduct the interview smoothly. In Rand's words, "[h]e felt himself rolling down a hill, without brakes. He could not stop." Matters escalate, as the passage quoted earlier shows, and Keating ends up not only failing to conceal his fear of Roark, but confessing it to boot.

### **Emotions and Cognition**

The Aristotelian conception of virtue and vice gives emotion a central role in their constitution. The emotions that partly constitute the virtues not only motivate right action, they also have cognitive power, insofar as they track what is truly valuable. Thus, the courageous person's confidence and fearlessness aid him in seeing which dangers are worth facing for which ends. By contrast, the emotions that partly constitute the vices track what is disvaluable, a spurious image of the good. Thus, the cowardly person's fearfulness and lack of confidence exaggerate the danger, becoming tools of distortion that distort or block the cognitive power of the intellect as well.

Clearly, Rand the novelist, like Aristotle the philosopher, sees the agent's emotional dispositions as a crucial component of his moral character, and as having the power to enhance or distort cognition (see II, 2 and 3 above). But what about Rand the philosopher? Rand's claim that "emotions are not tools of

cognition" (VOS, p. 29) has often been interpreted in a way that contradicts the picture she presents in her novels. However, this claim must be interpreted in the context of another important claim, viz., that "[e]motions are estimates of that which furthers man's values or threatens them, that which is *for* him or *against* him—lightning calculators giving him the sum of his profit or loss" (p. 27). The emotions of someone who wholeheartedly values the truly valuable—truth, reason et al.—will apprise her of what is truly good or bad in particular situations. Since emotions, unlike conscious reasoning, are "lightning" quick, without them she would often act too late or fail to act at all. Again, Rand would agree that since rational emotions, unlike deliberate, conscious reasoning, make available a vast store of evaluative knowledge, in the absence of such emotions a person would simply fail to see certain things. Without rational emotions, then, a person would make mistakes of judgment and act inappropriately or not at all. It is this vast store of knowledge embodied in her emotions that enables Dagny to recognize, "[i]n a single shock of emotion," that Ellis Wyatt's simple greeting signifies "forgiveness, understanding, acknowledgment" (AS, 157). And it is because Dagny knows that her emotions have cognitive power that she can surrender "her consciousness to a single sight and a single, wordless emotion . . . [aware] that what she now felt was the instantaneous total of the thoughts she did not have to name, the final sum of a long progression, like a voice telling her by means of a feeling" (674).

In her fiction, Rand also depicts the power of emotions to affect cognition in ways that are independent of the issue of virtue or vice. Moods and feelings induced by events in one's life, events to which they may be appropriate responses, can affect the way other things appear to one. In a couple of striking scenes in *Atlas Shrugged*, we see Hank Rearden first overcome by disgust at the world around him, a disgust that makes "the city seem sodden to him" (349), and then, on reaching Dagny's apartment, recover his sense of benevolence, a sense that enables him to see the city as a stupendous achievement of human creativity (351). In fact, the city is both sodden in some respects and a great achievement, but Rearden's disgust at the world hides its greatness till he has recovered the proper emotional state, a sense of benevolence.

As this discussion shows, some of Rand's stated views of the emotions, along with her depiction of them in her fiction, imply the view, so central to Aristotle's conception of virtue, that emotions have cognitive power. Hence, the claim that emotions are not tools of cognition must be interpreted to mean that they are not in themselves tools of cognition. Rather, they must be "programmed" by the intellect. As she states, "[m]an's emotional mechanism is like an electronic computer, which his mind has to program—and the programming consists of the values his mind chooses" (VOS, 28).

The idea that the emotions have to be programmed by the intellect, whereas the intellect can choose values independently of any help from the emotions, suggests a hierarchical relationship between intellect and emotion, and a unidirectional picture of moral and psychological development. First the intellect, functioning independently of the emotional faculty, collects the data and makes value-judgments; then it programs the emotional faculty. On this picture, the preprogrammed emotional faculty is inert, unable to make any value responses, and unable to play a fundamental role in forming or aiding the intellect.

However, if infants and young children (not to mention animals) have emotions in a pre-conceptual form—as they surely do—then emotions cannot be entirely dependent on the intellect. We feel fear, anger, contentment, empathy, and pleasure in a pre-conceptual form long before we acquire the capacity to make value-judgements. Insofar as these are responses to that which we sense as somehow good or bad for us, valuable or disvaluable, it follows that we are able to make value *responses* long before we are able to make value-*judgements*. Indeed, it is only because we have this pre-conceptual ability for responding to value that we can acquire the capacity for making value-judgments. Thus, pre-conceptual emotions are necessary for having any more than the most primitive values in the first place, and, thereby, for making value-judgments. Adult emotions build on these pre-conceptual emotions and the value-judgments they make possible. For example, adult fear typically contains not only the components of feeling and physiological response that a child's fear does, but also the value-judgment of the feared object as dangerous or threatening. Which objects are seen as fearful depends not on the judgments of an untouched intellect, but an intellect already shaped to some extent by our pre-conceptual emotions, and continually influenced by, even as it in turn influences, our adult emotions.<sup>15</sup>

Aristotle's picture of moral and psychological development as that of a process in which intellect and emotion grow and mature interdependently, each influencing the other, reflects these facts. It is, therefore, a more adequate account than Rand's hierarchical account of the emotions as programmed by an untouched intellect.

Rand's writings also often suggest that in a conflict between one's emotions and one's intellectual judgement, one must always opt for the latter, that the intellect is always more trustworthy than the emotions. However, we have already seen a counterexample to this in the scene where Dagny finds herself responding to Francisco happily, instead of with the intended coldness. One reason why one's emotional evaluation in a situation may be more trustworthy is that, as Rand herself points out, unlike the intellect, emotions can apprise us of a vast amount of evaluative knowledge. Given this, whether one should opt for the deliverances of one's emotions in a particular situation, or for one's intellectual judgement, depends on the general reliability of one's emotions vis-a-vis one's intellect in that sort of situation. The issue cannot be decided simply by appeal to a hierarchical relationship between intellect and emotion (even should this picture of a hierarchical relationship be correct). Indeed, some of the psychological nuances and complexities of Dominique's and Roark's relationships with Gail Wynand can be understood only as the result of each of them allowing their emotional responses to challenge their intellectual judgments. Consider the passage in which Dominique urges Wynand—the man who stands for everything she despises—to fire Ellsworth Toohey, because he is a threat to Wynand's beloved *Banner*—the paper that caters to everything she despises.

Gail, when I married you, I didn't know I'd come to feel this kind of loyalty to you. It contradicts everything I've done, it contradicts so much more than I can tell you—it's a sort of catastrophe for me, a turning point—don't ask me

why—it will take me years to understand—I know only that this is what I owe you (499-500).

She allows her feeling of loyalty to Wynand to dictate her action, even though she cannot quite understand why she feels this loyalty to him; she “knows” she “owes” him this warning, even though she cannot quite understand why she should want his paper saved. The fact that Wynand is an “innocent weapon” compared to Toohey, who is “a corrosive gas . . . the kind that eats lungs out” (500), neither justifies Dominique’s feeling of loyalty, nor supports her claim to “know” that she “owes” Wynand a warning. After all, even if Wynand is innocent compared to Toohey, his record of destruction can still only be classified as unambiguously evil. We can understand Dominique’s actions and words only if we interpret her as trusting her emotions to tell her something her intellect alone cannot yet grasp.

To reiterate: Rand’s depiction of virtuous individuals, and of the role of emotion in virtuous action, in her novels is closer to Aristotle’s views of these matters than her own stated views. But what about her conception of happiness, and of its relationship to virtue? It is to this question that I will now turn.

## HAPPINESS

### Rand’s Definition(s) of Happiness

(i) “Happiness is that state of consciousness which proceeds from the achievement of one’s values” (VOS, 28).

The values in question are rational values. “If you achieve that which is the good by a rational standard of value, it will necessarily make you happy; but that which makes you happy, by some undefined emotional standard, is not necessarily the good” (VOS, 29). (The implication of the second clause, that it is possible to be happy even if one’s values are irrational, is later taken back, so I will simply ignore it.)

(ii) “Happiness is a state of non-contradictory joy—a joy without penalty or guilt, a joy that does not clash with any of your values and does not work for your own destruction . . . .” Rand continues: “[h]appiness is possible only to a rational man, the man who desires nothing but rational goals, seeks nothing but rational values and finds his joy in nothing but rational actions” (29).

These definitions make two important points:

- (1) Happiness is a state of consciousness.
- (2) It is a positive, harmonious (non-contradictory) state of consciousness that results from the achievement of one’s rational values, and only from such values.

Rand also gives a less mentalistic definition of happiness in VOS.

(iii) “Happiness is the successful state of life” (27). More fully, “[t]he maintenance of life and the pursuit of happiness are not two separate issues . . . [but] two aspects of the same

achievement. Existentially, the activity of pursuing rational goals is the activity of maintaining one's life; psychologically, its result, reward and concomitant is an emotional state of happiness" (29).

Putting these thoughts together, we can say that, for Rand, happiness is a successful state of life, and the positive state of consciousness that accompanies and results from such a life.

The values Rand has in mind when she says that happiness results from the achievement of one's rational values are existential or external values or life-goals, most importantly, career and romantic love. Thus, when she says that *The Fountainhead* and *Atlas Shrugged* have happy endings, and *We the Living* a tragic ending, she means that *The Fountainhead* and *Atlas Shrugged* end with the success of her protagonists in achieving their most important life-goals through rational (moral) action: succeeding at least in their careers, but also, in the case of the most important characters, in their romantic attachments. And when she says that *We the Living* has a tragic ending, she means that her protagonists fail to achieve their most important life-goals. This is, of course, in keeping with the usual understanding of a happy or tragic ending. What is noteworthy is that Rand gives no hint that she regards spiritual success in the absence of existential success—i.e., success in remaining true to one's rational values in the face of existential failure—as constituting any part of happiness. If she had, then she would have acknowledged that *We the Living* was not entirely a tragedy. After all, in this novel only Leo is destroyed spiritually; Kira triumphs, and Andrei learns the meaning of love and individuality. Elsewhere, too, as we have seen, Rand equates a happy or successful life with a life in which we achieve our rational external values through virtuous action.

On this conception of happiness—the conception standardly accepted in interpretations of her views—virtue is only a means to happiness. Yet many of her claims—as also her portrayal of her characters—imply a different view, the view that a life in which we fail to achieve our most important external values, but still continue to act honestly, justly, and with integrity, is also to some extent a successful and, therefore, happy life. In other words, many of her claims imply the Aristotelian view that a virtuous life is partly *constitutive* of a happy life. The most important texts supporting this view are the ones that deal with Rand's conception of the most important values and their connection to happiness.

As we have already seen, the cardinal values, the values that are expressed by the cardinal virtues, are the largely psychological or "internal" higher-order values of reason, purpose, and self-esteem. As she makes clear, to truly value reason is to have a commitment to living rationally, and (presumably) to derive pleasure from living rationally. Likewise, to truly value having a purpose is to have a commitment to living a life of productive activity, and to derive pleasure from living productively. The cardinal virtues of rationality and of productiveness, then, are exercised in rational and productive activity that is motivated in this wholehearted way by the value of reason and purposiveness. And it is in a life characterized by the virtues of rationality and productivity that one maintains and expresses love of reason and purposiveness. The third cardinal

value, a sense of self-esteem, is the sense of oneself as being *able* to achieve happiness and being worthy of happiness (AS, 936, FNI, 128). Hence, someone who truly values self-esteem will continually strive to become—and remain—the sort of person who is both capable of happiness and *worthy* of happiness. The virtue necessary for self-esteem, says Rand, is the virtue of pride or *moral ambitiousness*, the virtue of acting to achieve one's own moral perfection (VOS, 27). Only by acting to perfect ourselves can we achieve and maintain self-esteem, and only by valuing self-esteem will we be motivated to act with pride. It is in a life characterized by the virtue of pride, then, that one expresses the value one places on self-esteem.

It follows that, so long as one can act virtuously, one is guaranteed success at achieving or maintaining the three supreme values—reason, purpose, and self-esteem—regardless of success or failure in achieving one's external values. So, if happiness is a successful state of life, then such "inner" success must count not only as a necessary *means* to happiness, but as itself a major *part* of happiness. I will refer to the life of merely inner success as a life of partial or "inner" happiness, and the life of both inner and outer success as full happiness. Images of both partial and full happiness occur in several passages in Rand's novels.

### **Images of Happiness in Rand's Novels**

#### (i) Partial happiness

Roark in the quarry (*The Fountainhead*): Roark's months in the quarry are shot through with pain—pain at the loss of the opportunity to be doing the thing he loves. Yet he is not entirely unhappy. His consciousness of having done the right thing in refusing to build buildings that violate his architectural principles, and his sense of purpose in being engaged in a "clean," worthwhile activity in the quarry, impart to his life a certain serenity and quiet satisfaction that are part of happiness.

Francisco after he has given up Dagny and his work, and decided to assume a new persona for the public (*Atlas Shrugged*): After his initial tortured struggle, when he begs Dagny to help him to refuse John Galt's call to "strike" and to stay with her, Francisco achieves a measure of serenity in the knowledge that his renunciation of his life-goals is necessary for a deeper and longer-lasting success. His house in Galt's Gulch serves as a splendid metaphor for his state of mind in those years of painful renunciation: the "silent, locked exterior" of the house bespeaks sorrow and loneliness—the interior is filled with an "invigorating brightness" (AS, 710).

Interestingly, Peikoff also draws on these facts about the psychology of Rand's heroes to come to the conclusion that "[v]irtue does ensure happiness in a certain sense, just as it ensures practicality" (*Objectivism*, 339), "not the full happiness of having achieved one's values in reality, but the premonitory radiance of knowing that such achievement is possible" (340). Indeed, Peikoff puts it even more strongly—and somewhat misleadingly—when he says that someone like Roark is "a happy person even when living through an unhappy period" (339-400). He distinguishes between the achievement of existential and philosophical values, and between full happiness and "happiness in a certain

sense,” or “metaphysical pleasure” (340). Yet he denies that “the achievement of philosophical values,” reason, purpose, and self-esteem, which we achieve and maintain only through virtue, constitutes a form of success, describing it instead as the achievement of “the ability to succeed.” However, if achieving and maintaining the cardinal values and virtue is not a form of success, but success is necessary to happiness, then it is hard to see how virtue can “ensure happiness” in any sense of the word. To consistently maintain the thesis that virtue ensures happiness “in a certain sense,” Peikoff would have to reject the canonical view that equates happiness with the state of consciousness that results from existential success, and sees virtue as entirely a means to happiness. But this should not be a problem. For, as we have seen, some of Rand’s own theoretical statements imply the rejection of the canonical view, and her fiction constitutes a powerful argument in support of this rejection.

### **(ii) Full happiness**

Dagny and Francisco in the early days of their relationship, before he (apparently) turns into a playboy and their relationship comes to an end: The description of her state of mind after her first sexual encounter with Francisco is a good example of full happiness. “[W]hen she thought that she would not sleep . . . her last thought was of the times when she had wanted to express, but found no way to do it, an instant’s knowledge of a feeling greater than happiness, the feeling of one’s blessing upon the whole of the earth, the feeling of being in love with the fact that one exists and in this kind of world” (AS, 105-6).

Dominique and Roark after they are united and he has become a successful architect. The passage that captures her happiness best, however, occurs shortly before this, when she decides to leave Wynand and go back to Roark—and the world she has rejected out of fear and disgust.

Dominique lay stretched out on the shore of the lake . . . Flat on her back, hands crossed under her head, she studied the motion of leaves against the sky. It was an earnest occupation, giving her full contentment. She thought, it’s a lovely kind of green . . . The fire around the edges is the sun . . . The spots of light weaving in circles—that’s the lake . . . the lake is beautiful today . . . I have never been able to enjoy it before, the sight of the earth . . . I thought of those who owned it and then it hurt me too much. I can love it now. They don’t own it . . . The earth is beautiful . . . (FH, 665-66).

She thought, I’ve learned to bear anything except happiness. I must learn how to carry it. How not to break under it (666).

## CONCLUSION

We have seen that Rand's views about the three supreme values and the virtues required for them leads to the view that virtuous activity is itself partly constitutive of happiness. For on this view virtuous activity is both a means to, and an expression or realization of, the three supreme values—reason, purpose, and self-esteem—and these values are both the means to, and the realization of, one's ultimate value, happiness. More formally:

1. Virtuous activity is both a means to, and realization of, the supreme values.
2. These values are both a means to, and realization of, an important part of happiness.
3. Hence, virtuous activity is also both a means to, and realization of, an important part of happiness.

Virtuous activity is inherently deeply satisfying or happiness-making. That is, the satisfaction that comes from virtuous activity is "embedded" in it the way the pleasure that comes from walking along the beach is embedded in the activity. The passages from Rand's novels discussed above show why this is so. In acting virtuously and, thereby, aiming at, and expressing, our values, we actualize a clear-sighted view of our selves and of external reality. A virtuous life thus brings with it a sense of harmony and of freedom—a justified sense of efficacy, of the power of one's agency to deal with external obstacles. It is this sort of enduring reality-oriented pleasure and deep satisfaction that is an essential and central part of happiness. It is only when we cease to act virtuously that we lose happiness altogether. Henry Cameron and Steven Mallory, minor characters in *The Fountainhead*, are examples of individuals who allow their existential failures to damage their inner resources, their capacity for virtuous action. When first introduced to the reader, they are shown as bitter, self-destructive individuals, who are rescued from this state only with Roark's help and kindness. It is also, of course, possible to never develop one's inner resources and, therefore, to never achieve happiness. Keating is a case in point.

Insofar as the virtues are a constitutive part of happiness, they are ends in themselves. But they are also, of course, means to happiness. As traits and acts that put us in the best state to achieve and maintain a successful state of life, they aim at bringing about certain states of affairs in the world. For example, the aim of being just is to bring about just states of affairs. But success in doing this often depends on circumstances that are independent of the agent's actions. Thus, the success of a judge in acquitting an innocent defendant depends not only on his acting justly himself, but also on the others involved acting justly and efficiently, as well as on luck in gathering the evidence. In short, because virtuous action is a means to external success, and because external success is essential to full happiness, virtuous action is also a means to happiness.

It is because it has this instrumental relationship to happiness that virtue is never sufficient for full happiness. For it is possible to act virtuously, yet fail, through misfortune, to achieve one's most important goals. Such a life, though (necessarily) not unhappy, is not fully happy either. An unqualifiedly happy life is one in which one's actions are largely rewarded by success, and one's sense of satisfaction in one's life is partly derived from this success.

## NOTES:

1. This paper was first written for a talk at the Institute for Objectivist Studies Summer Seminar in July 1997. I would like to thank the IOS for the opportunity to present it, Lester Hunt, David Kelley, and Chris Swoyer for their helpful written comments, and Larry Abrams and Murray Franck for taking the time to read it and discuss it with me.
2. Ayn Rand, *Atlas Shrugged* (AS), 35th Anniversary Edition (New York: Signet, 1957), 939, *For the New Intellectual* (FNI) (New York: Signet), 131.
3. Leonard Peikoff, *Objectivism: The Philosophy of Ayn Rand* (New York: Meridian, 1991), 328. See also David Kelley, *Unrugged Individualism: The Selfish Basis of Benevolence* (New York: Institute for Objectivist Studies, 1996), for an entirely instrumentalist defense of the virtue of benevolence.
4. In "Flourishing Egoism," Lester Hunt takes a stronger position, arguing that in Rand's ethics the virtues are constitutive of happiness, or "an essential part of it" (*Social Philosophy and Policy*, eds. E.F. Paul, F. Miller, and J. Paul, Vol. 16, 1, Winter 1999), 93. My view, as stated above, is that although this idea is implicit in her writings, her official view is different. Hunt and I also develop very different (though largely compatible) lines of argument for the claim that the constitutive view is to be found in Rand. Whereas there are some other philosophers who would agree with this claim as part of a broadly Aristotelian interpretation of Rand's ethics (see, for example, Tibor Machan, *Ayn Rand* [New York: Peter Lang, 1999]), no one else actually argues for it.
5. AS, 930, "The Objectivist Ethics," *The Virtue of Selfishness* (VOS) (New York: Signet, 1961), 27.
6. More precisely, it is a means in normal circumstances. An act of resistance to tyranny in the cause of justice, but without the expectation of furthering this cause, is only an expression of certain values, not a means to any.
7. I discuss Rand's conception of happiness below, pp. 44-46.
8. This interpretation is also supported by these words of Francisco to Dagny: "[e]very form of happiness is one, every desire is driven by the same motor—by our love for a single value, for the highest potentiality of our own existence—and every achievement is an expression of it" (AS, 708). I take this to mean that every achievement is an expression of our love for, and desire to realize, our highest potential, and happiness consists of doing so. The idea that happiness lies in self-realization is not often (if at all) voiced in Rand's philosophical writings, but it is often illustrated in her novels. Thanks to David Kelley for bringing this passage to my attention.
9. There is a reference to character in the definition of the virtue of pride (AS, 938, FNI, 130), but here, too, character is that which pride aims to build; pride itself is not said to be a character trait.
10. The problem goes deeper than this, for it affects not only moral response, but all belief and all action. As Ronald de Sousa argues in *The Rationality of Emotion* (Cambridge: MIT Press, 1987), no amount of knowledge or degree of logical and inductive acumen would enable us to solve the "frame problem," i.e., the problem of knowing what is relevant, what irrelevant, in the mass of data on the basis of which we make inferences, form beliefs, and decide on action (192-96). To solve the frame problem, we need emotions. In *Descartes' Error: Emotion, Reason and the Human Brain* (New York: Grosset/Putnam, 1994), the neurologist Antonio Damasio discusses patients with high IQs but impaired emotional faculties who are unable to make the simplest of decisions. One patient is unable to choose an appointment time with Damasio because he can give countless arguments for and against any time that Damasio proposes. According to Damasio, such people have lost their "somatic markers," or gut feelings, that tell us that certain courses of action are good for us, others bad. See Daniel Goleman, *Emotional Intelligence* (EI) (New York: Bantam, 1995), 27-29, 52-54 for an excellent discussion of Damasio's work as well as of the neurological, psychological, and biological research on the emotions.
11. Hervey Cleckley, *The Mask of Sanity* (St. Louis: C.V. Mosby, 1964), argues that the psychopath's failure to be a moral agent is due to the fact that his stunted emotional capacity renders him incapable of seeing the significance of things.
12. Aristotle, *Nicomachean Ethics* (NE), tr. Terence I. Irwin (Indianapolis: Hackett Publishing Co., 1985), 1106b36-1107a2.
13. *The Fountainhead* (FH), 50th Anniversary Edition (New York: Signet, 1971), 191-92.
14. There is a suggestion in the text that Rand thinks that the city is seen veridically only when it is seen as an achievement, and not also as sodden. But my interpretation is both compatible with the text and more accurate as a description of the city.
15. Some of the best evidence for the importance of normal emotional development to normal intellectual development comes from studies of autism. Autistic people show an inability to understand complex personal relationships and emotional nuances due to neurological abnormalities that prevent normal emotional development. See Temple Grandin's autobiographical account in *Thinking in Pictures* (New York: Vintage Books, 1995), especially ch. 4, "Learning Empathy."
16. Elsewhere I have also argued that virtuous activities and attitudes are partly constitutive of certain external goods, such as friendship (see my "Friendship, Justice and Supererogation," *American Philosophical Quarterly*, V. 22, 2, April 1985, pp. 123-131.) Thus, the virtues of benevolence and of justice are partly constitutive of the best (i.e., most enduring, meaningful, and fulfilling) friendships. If friendship is a means to happiness, then it follows that virtue is a means to happiness in yet another way.

# Foundations, Philosophy, and The Location of Socrates' Feet.

Nino Langiulli

The following discussion attempts to make several interrelated points. The first is to offer an interpretation of a line in Plato's *Phaedo* as a response to Aristophanes' conservative view, i.e., philosophy undermines traditional views of reality and, *a fortiori*, of morality, politics, and esthetics as well. The second is that the line is doubly ironic such that the second turn of the irony is Plato's reply to the ancient liberal view as represented by the Sophists—the view that reality, morality, and esthetics are solely matters of opinion and politics solely a matter of power since transcendent points of reference are impossible. Protagoras' dictum that “man is the measure: of the things that are that they are; of the things that are not that they are not” means that only immanence is possible. This is to say more specifically that truth and falsity cannot be known only opinions possessed; goodness and badness cannot be achieved only pleasures and pains can be; unity and sameness do not exist only multiplicity and difference; justice and injustice are not possible only the interests of the strong and the weak; beauty and ugliness cannot be discovered only likes and dislikes.

The third point is a reply to the contemporary sophists, i.e., the anti-foundationalists or contemporary liberals. They, in order to affirm humanity and the human things, maintain that foundations do not exist and philosophy, as the attempt to uncover them, is either futile at worst or poetry at best.

The argument that foundations do not exist and that philosophy is futile is self-contradictory. The argument that philosophy is just poetry, and bad poetry at that, is Nietzsche's view. But, unlike his liberal epigoni who are positive and optimistic about humanity, Nietzsche follows his argument of philosophy as poetry to its logical conclusion—to a negative and pessimistic humanism.

One important source for his argument is his interpretation of the *Phaedo*—an argument and interpretation with which those offered herein are in conflict.<sup>1</sup> That conflict will be treated at the end of this discussion wherein the true understanding and limits of humanity are disclosed most radically and most fundamentally in philosophy's search for a foundation.

In a remark near the beginning of the *Phaedo* (61d) that is arresting and slyly cryptic, Plato seems to make a joke about philosophy. Given its immediate context, it has an evident and not especially profound meaning. But it is its very obviousness that naggingly draws the curious reader beyond its surface

triviality into the theme and core of the dialogue, namely, the relationship between philosophy, the soul, and mortality. What is more, the sentence recollects others in the *Apology*. It may be that Plato is having a joke at the expense of Socrates, or even more interestingly, employing his irrepressible irony to make a joke about philosophy itself.<sup>2</sup> But the point—the connection, the joke, the irony—cannot be grasped unless and until the reader understands that the sentence is a response to Aristophanes' portrayal of Socrates in *The Clouds*—a portrayal addressed by Plato in the *Apology*. In the play Socrates is the character who, suspended in a basket in the air, represents for Aristophanes, the corruption of Athenian religion and morality by means of scientific speculation and sophistry. This is to say that the gods are replaced by natural phenomena in the explanation of things and that traditional morality is challenged and undermined by clever, subtle though specious arguments. The young man, Pheidippides, by the end of the play denies the existence of the gods, beats his father, Strepsiades, and argues, moreover, that his parents deserve to be beaten.

In the *Apology*, Plato treats the character of Socrates in *The Clouds* as a caricature and alludes to it as his own Socratic character gives his first speech before his Athenian judges recalling the underlying bias against him.

Socrates is guilty of criminal meddling, in that he inquires into things below the earth and in the sky, and makes the weaker argument defeat the stronger and teaches others to follow his example. It runs something like that. You have seen it for yourselves in the play by Aristophanes, where Socrates goes whirling round, proclaiming that he is walking on air, and uttering a great deal of nonsense about things of which I know nothing whatsoever (19 b-c).

This caricature passing for a description is, as the character Socrates himself says (*Apology*, 23d), the stock charge against all philosophers. They do not have their feet on the ground because their heads are in the clouds, asking questions about those matters which everyone else takes for granted or about things that are beyond ordinary human experience and thereby not worth knowing. They challenge the common-sense opinions in morals and politics, as well as the common opinions in religion, but wish at the same time to stand above the beyond the practical consequences of doing so. By means of these questions and challenges, they attract those young people who are eager for novelty and innovation and at the same time contemptuous of tired old views and manners, teaching them rhetorical and dialectical tricks which turn them into glib, not to speak of snide, smart-alecks ready to jettison God, family, friends, and country. Philosophers are, at best, impractical muddle-headed fools or, at worst, careless corrupters of youth. The philosopher's own distinctions between sophists, wise men, and philosophers are too subtle or arch to have any practical value or count very much when bodies are being count-

ed for the showdown. Philosophers make no tangible contribution to human life. Professionals, such as physicians, lawyers, and teachers do. So do priests, politicians, and artisans, as do persons engaged in agriculture and business. Poets and other artists may go unappreciated but the marks they leave are among the most characteristic and most lasting of cultural endeavors. Even those other theoreticians or academics such as scientists, historians and mathematicians perform tasks which can be put to some good use in the human enterprise. But philosophers, committed as they are, to the most theoretical of activities and therefore the most practically useless, should best be ignored and forgotten. Two of them, Richard Rorty and Jacques Derrida, have come in recent years to recognize and acknowledge this, despite their nostalgia for philosophy's past.

The tone of this expansion of the caricature may lead the reader to believe that the caricature should be dismissed. It *may* have been misapplied to Socrates, as Plato suggests, but sound political thinking requires that it not be dismissed altogether. It might not apply to the "genuine" philosopher who makes much of the difference between theoretical knowledge and practical wisdom and who, much more pertinently, *knows* not only the importance of ideas but the *danger* of them and cares about the danger. Now there is a kind of dealer in ideas, that obnoxious brand of human animal whom Eva Brann has shrewdly denominated a "monster,"<sup>4</sup> quite a while after Julien Benda's attack but a few years before Paul Johnson's more recent admonition,<sup>5</sup> namely, the intellectual. It is a feature of his monstrosity that he may frequently pass as a philosopher, but one who righteously wants to affect or change the world. It just might be that it is precisely this person whom Aristophanes portrays as the dangler from a basket. And who else could serve better the purposes of the play than that well known "pest," Socrates. The *Apology* is not nescient of the lack of ease in distinguishing the philosopher from the sophists (the ancient intellectuals). The character, Socrates, in his own defense, appeals to that which his judges, and most everyone else, can understand as a distinguishing mark: he charges no fee for his ideas. The fact that he has no "doctrine" is another distinguishing mark, but this is part of what is not so easy to understand.

In the *Phaedo*, however, Plato offers his reply to the caricature in that cryptic and seemingly negligible remark which is both the subject of this essay and the rubric for the theme of the dialogue—the relationship between philosophy, the soul, and mortality.

And as he spoke he lowered his feet to the ground and sat like this for the rest of the discussion (61d).

Not only does Socrates *not* have his feet dangling in mid-air, he has them planted squarely on the ground. The discussion in the *Phaedo* is, therefore, concerned with the relationship of philosophy to human mortality. It is about what one does in the face of death (not only one's own but those of loved ones, a more significant trial) and what is the best kind of life, granting the ineluctabil-

ity of death. This is no abstract and theoretical discussion for Socrates, inasmuch as he must suffer his own death shortly. As the *Phaedo* says:

At any rate, said Socrates, I hardly think that anyone who hears us now—even a comic poet—would say that I am wasting time and discoursing on subjects which do not concern me (69c).

This remark is corroborated by Plato when, speaking of arguments in general and of the arguments for immortality in particular, he has Socrates say:

We must not let it enter our minds that there may be no validity in argument. On the contrary, we should recognize that we ourselves are still intellectual invalids, but that we must brace ourselves and do our best to become healthy—you and the others partly with a view to the rest of your lives, but I directly in view of my death, because at the moment I am in danger of regarding it not philosophically but self-assertively. You know how, in an argument, people who have no real education care nothing for the facts of the case, and are only anxious to get their point of view accepted by the audience? Well, I feel that at this moment, that my anxiety will not be to convince my audience except incidentally, but to produce the strongest possible conviction in myself (90E-91B).

Socrates has his feet on the ground; he is discussing a question of the utmost relevance to himself and he is deadly serious about it. It is a discussion of mortality and immortality in the face of his own execution.

The reader is then struck with a marvelously contorted and disconcerting irony. That he has come to expect such a thing from the masters of irony (Socrates and more so Plato) is only meager insulation when a new one blows in—especially if it is a compound irony. The point here must be made explicitly and it must be made clearly and distinctly. Although it is a mark of great writing to be subtle and even arch in exposition, it is not a mark of great teaching. Thomas Aquinas says of Plato, for example, that “habuit malum modum docendi.”<sup>6</sup> For the sake of teaching, therefore, and at the risk of unsophistication, the irony can be untangled in the following manner.

Socrates has been accused of being an impractical theoretician with his feet off the ground, who also undermines with sophistry the common opinions on which civil society rests. But in the *Apology*, and in historical fact as well, according to Aristotle,<sup>7</sup> Socrates is particularly concerned with the human things, the domain of morals and politics—the earthly things—the things that

are his business. As the Athenian “gadfly,” he is seeing to it that those conditions which make philosophy and indeed *human* life possible are pursued and preserved,

. . . for I spend all my time going about trying to persuade you, young and old, to make your first and chief concern not for your bodies nor for your possessions, but for the highest welfare of your soul, proclaiming as I go ‘Wealth does not bring goodness, but goodness brings wealth and every other blessing both to the individual and to the State . . . .’

It is literally true (even if it sounds rather comical) that God has specially appointed me to this city, as though it were a large thoroughbred horse, which because of its great size, is inclined to be lazy and needs the stimulation of some stringing fly. It seems to me God has attached me to this city to perform the office of such a fly (*Apology*, 30b; 30e).

Pest though he may be, he is bothering about the human things and to that extent has his feet on the ground.

Now, in the *Phaedo*, to repeat the point made earlier, Plato responds to the stock caricature by asserting that Socrates, in his discourse about the soul and mortality in the face of his own death, has his feet on the ground, keeping them there for the entire length of the discussion. But the irony of the matter is that *that* very discussion about the soul and the things which it is like—the “forms” or “ideas”—is about things which are, of course, not on the earth nor of the earth but “above” it. The discussion turns out to be not about the “human things” but about the “divine” things—not about the physical, moral or political realms but about the metaphysical realm.

If all these absolute realities, such as beauty and goodness, which we are always talking about, really exist; if it is to them as we rediscover our own former knowledge of them, that we refer, as copies to their patterns, all the objects of our physical perception—if these realities exist, does it not follow that our souls must exist too even before our birth whereas if they do not exist, our discussion would seem a waste of time (76e).

Then let us return to the same examples which we were discussing before. Does that absolute

reality which we define in our discussion remain always constant and invariable, or not? Does absolute equality or beauty or any other independent entity which really exists ever admit change of any kind? Or does each one of these uniform and independent entities remain always constant and invariable, never admitting any alteration in any respect or in any sense? (77d).

But when it [the soul] investigates by itself, it passes into the realm of the pure and immortal and changeless, and being of a kindred nature, when it is once independent and free from interference, consorts with it always and strays no longer, but remains, in that realm of the absolute, constant and invariable, through contact with beings of a similar nature. And this condition of the soul we call wisdom (79d).

Now, Cebes, he said, see whether this is our conclusion from all that we have said. The soul is most like that which is divine, immortal, intelligible, uniform, indissoluble, and ever self-consistent and invariable, whereas body is most like that which is human, mortal, multiform, unintelligible, dissoluble, and never self-consistent (80b).

It would seem that the discussion places Socrates' feet not only in the clouds, but "above" and "beyond" the clouds. Is it perhaps that Plato is *simply* confirming Aristophanes' caricature and thereby confirming the case not only against Socrates but also against philosophy itself—indeed against "foundational" or "onto-theological" philosophy as the current fashion will have it? And if this be so, then perhaps too we can understand another odd comment from the very writer and chief speaker of the *Phaedo*: "Plato was ill" (59b) and not present for the discussion, suggesting his separation not only from "dubious" arguments for immortality but also from high-flying philosophy.<sup>8</sup> Perhaps, but not so "simply."

On the one hand a discussion of mortality, immortality, and the nature of the soul, would be perfectly appropriate for someone who is about to die. Such concerns in these circumstances are not the exclusive domain of philosophers. The might even be regarded as a mark of "realism" and "practicality"—of having one's feet on the ground. Yet the discussion is unmistakably about metaphysical things, i.e., the "divine" and not the "human." Hence the platonic irony. But there is another twist to it—the *final* twist. Just as with other things human, the irony cannot be endless, or else it could not be, or be understood.

If it be so that there exists a polarity and a tension between politics and phi-

losophy, as the *Apology*<sup>9</sup> and other dialogues of Plato affirm, and if it be so that the proper business of philosophers is the inquiry into and investigation of everything and anything unto their ultimate grounds, even if the inquiry and investigation should conclude that no such grounds exist or that the very inquiry itself is futile or misplaced, then it follows that when the philosopher is concerned with metaphysics, he does have *his* feet on the ground; he is minding *his* own business. We must say here, again with our attention on the concerns and arguments of current anti-foundationalists, e.g., the later Heidegger, Derrida, Rorty and their followers, that the constitutive element of metaphysics is the *inquiry* itself. This is true of every period of history of philosophy, whether it be the ancient, medieval, modern or contemporary period. It also matters not whether the priority is ontological or epistemological. For to begin with one is to reach the other. The even halfway serious posing of the question that asks whether or not there is a primary or fundamental sense of “reality” is deep into foundations. The particular *answer* that is proposed by any given investigation is of secondary importance in determining whether or not a foundational endeavor takes place. The fact alone, of the inquiry seriously begun, is sufficient. The inquirer is most faithfully the philosopher when he takes any question to its end to find out whether or not the answer satisfies the question.

Although the philosopher must consider the nature of political, poetic or technical things, he must leave politics to the politicians, poetry to the poets, and technical things to the technicians. This not to say, however, that because of his deflection from the careers of politician, poet or technician, he could not or ought not “advise,” “guide” or “teach” them. It is precisely because it is in his business to investigate these things, and to distinguish or differentiate between them in the course of thinking about their natures that he can advise, guide, or teach.

Herein lies the *telos*, the end, of philosophy, a primal theme which Plato illustrates in such dialogues as the *Parmenides* and the *Sophist* on the one hand and the *Apology* and *Phaedo* on the other. A point throughout is the “wisdom” of Socrates, the only wisdom to which he admits—most explicitly in the *Apology* (21d; 23a-b) but operatively elsewhere. That wisdom, as Socrates describes it paradoxically, is a knowledge of his own ignorance. This means an understanding of the limits of his own knowledge (and by extension, of human knowledge as such), an understanding achieved through sustained inquiry into the differences among opinions about things (and by extension, among things); thus the Socratic dialogue.

The tensions between politics and philosophy lurk in every corner of the *Apology* along with the crisscrossing contrasts among politics, poetry, “*techne*” (the artisans) and philosophy—their activities, ends, and presumptions. The politicians’ claim to know what they do not know (*Apology*, 22a); the poets speak true things but by virtue of a divine inspiration and not from knowledge (*Apology*, 22b-c); and the artisans, who do know (as the “know how” of a craft) presume to speak on everything (*Apology*, 22d). But the philosopher, in know-

ing the limits of politics, poetry, and “technology,” as well as of philosophy, has, as it were, clipped his own wings, not leaving the ground of what he knows.

Plato speaks equally dramatically of the dangers of politics to philosophy in relating the risks to Socrates both from the democracy (*Apology*, 32a-c) and from the oligarchy (*Apology*, 32c-e). The danger from the oligarchy is particularly poignant inasmuch as its leaders had been “students” of Socrates.

The corruption of a philosopher’s teaching in the hands of political activists, even among the most “faithful” of them is not an uncommon complaint in the history of the tension between the actors and the ideologues. Most recently, what might be called “Marxian orthodoxy” has disengaged itself from the failed practices of the activist disciples (Lenin, Trotsky, Stalin, or Mao) by means of the distinction between pure and not yet applied Marxism on one hand and applied (e.g., in the Soviet Union) but hardly pure pseudo-Marxism on the other.

The complaint is self-serving and even puzzling in the face of Marx’s own insistence in the *Theses on Feuerbach* (1845) that “the goal now is not to interpret but to change the world.” The purists’ interpretation of the changes which were imposed on the world (i.e., the several totalitarian and repressive socialist and communist regimes) is that, because they *were* so universally infelicitous, they could not have been “authentic.” Non-Marxist, even anti-Marxist adulterations must have been allowed to creep in.

Change has been integral to the agenda of much of the “modern project,” from Descartes through the Enlightenment to Marx and Nietzsche—change in thought directed toward change in nature. The aim has been to change (or at least to dominate) nature, in the physical as well as in the moral realms; and the effort has not been unsuccessful—but at the cost of “authenticity.”

Or, to put it another way, how much and what kinds of change are compatible with “authenticity”? Is the goal perhaps the actualization in this world of a New Jerusalem, completely secular, of course, and a monument to Marxist (or some other solely human) purity? A lesson in “feet upon the ground” might well be in order at this point.

Plato himself knew the temptation—and the price—of trying to make of philosophy more than a limit to thought, the enticements of a chance at last to really guide a ruler. His “Sicilian fling” or misadventure, his attempt to teach the tyrant Dionysius some philosophy in order that he might govern well, was a bitter lesson for Plato (*Seventh Letter*, 345c-352a). It is not the role of philosophy to instruct princes or to plan cities.

For philosophy to be of any use to would-be politicians, poets, or “technicians,” it is necessary for them first to desire to understand the limits of knowledge, the finitude of human understanding. Philosophy begins with the desire (a kind of *eros*) to know, but it ends as a love (thus the *philo*) of understanding,

that is, as the ability to differentiate the natures and limits of things and opinions about them.

When the desire to know is clipped short, when the range of possible differentiations is pre-emptively hemmed in, to make it more accessible, then we get the stingy “understanding” found among contemporary literary theorists. Making use of the fashionable “method” of differentiating, that is, “deconstruction”—they arrive at no more profound “hegemonic structures” than race, sex or gender, and class for their universal categories or ultimate explanations. Unlike the poets in the *Apology*, these literati are hardly divinely inspired.

The “artisans” of our time are just as puffed up and unaware of the moderating role of philosophy as were their counterparts in the *Apology*. Technology is a kind of knowledge and it is becoming ever more powerful; but it is not omniscience. The minds of technologists are filled with theories. Philosophy, as the criticism of theories, is the killer of their joy in ideas. On the one hand we have those pelagian optimists (for example, concerning environmental problems) who believe that more technology is the solution to problems brought about by technology—indeed scoffing at the very suggestion of human imperfectability. And, on the other hand (again in the context of environmental problems) we have the gnostic morticians who would “cut and run” before the battle is half-begun, in the pseudo-humble belief that “Nature” would be better off without the human species. In the first case we have the careless arrogance of conquerors of nature; in the second we have the presumptuous quietism of abject submission to it. The moderate voice of reason, with which philosophy speaks, is neither the voice of God nor the voice of doom.

The philosopher does have his feet on the ground, when among his many and sundry conversations, he discusses the “divine” things; for in so doing he profoundly understands his humanity. His search for God or the gods is an indication that he is not God or one of the gods. In understanding his humanity, he recognizes and acknowledges his mortality. The inquiry into the immortality of the “Soul,” which like the “Forms” is universal and public, uncovers the radical “mortality of the individual self.”<sup>11</sup>

In the act of distinguishing the human things from the divine—the activity of making distinctions and therefore recognizing differences being its peculiar and specific work—philosophy resists the most tempting and dangerous of foundationalisms. I speak of that naked tempter—the individual ego as the sole and ultimate arbiter of what is true and what is right and even what is pleasing.

With his feet on the ground and his mind on the “divine” or metaphysical things (dare we say on the “grounds” of things), Socrates is portrayed by Plato as the embodiment and coincidence of philosophy, humanity, and mortality.<sup>12</sup> In other words, it is in the act of philosophizing<sup>13</sup> that both the humanity and the finitude of human beings is discussed in the most *radical* way.<sup>14</sup>

**References**

- Aquinas, Thomas. *In Aristotelis Librum De Anima Commentarium*. 3rd edition. Ed. A.M. Pirotta. Turin: Marietti, 1948.
- Aristophanes. *The Clouds in Five Comedies*. Tr. B.B. Rogers. New York: Doubleday, 1955.
- Aristotle. *The Basic Works*. Ed. Richard McKeon. New York: Random House, 1941.
- Brann, Eva. "Commencement Address." *St. John's College Magazine*, Annapolis, July, 1974.
- Hartle, Ann. *Death and the Disinterested Spectator*. Albany: State University of New York Press, 1986.
- Heidegger, Martin. "Discussion Between Martin Heidegger and Ernst Cassirer." Tr. Francis Slade. *The Existentialist Tradition*. Ed. Nino Langiulli. New York: Doubleday, 1971, reprinted as *European Existentialism* with a new introduction. New Brunswick: Transaction Books, 1997.
- Johnson, Paul. *Intellectuals*. New York: Harper & Row, 1989.
- Kierkegaard, Soren. *The Concept of Irony*. Tr. L.M. Capel. Bloomington: Indiana U. Press, 1965.
- Nietzsche, Friedrich. *The Birth of Tragedy in The Birth of Tragedy and the Genealogy of Morals*. Tr. R. Golfing. New York: Doubleday, 1956.
- Plato. *The Collected Dialogues*. Ed. E. Hamilton & H. Cairns. New York: Pantheon (Random House), 1961.
- Stern, Paul. *Socratic Rationalism and Political Philosophy: An Interpretation of Plato's Phaedo*. Albany: State University of New York Press, 1993.

## NOTES

1. A fairly recent book on the *Phaedo*—Paul Stern's *Socratic Rationalism and Political Philosophy: An Interpretation of Plato's Phaedo* (Albany: State University of New York Press, 1993)—covers some of the same ground and even shares one of the conclusions reached here. Stern, however, has a different understanding of Plato's notion of the foundation for philosophy. Moreover, the line and its interpretation which are the point of departure for this discussion are not treated in the book nor is Plato's pun seen as a more profound response to Aristophanes than the one given in the *Apology*. Thus Stern misses the double irony as the vehicle whereby Plato indicates his deepest understanding of both political activity and political philosophy. Nevertheless on page 145 of his book, Stern reaches the correct conclusion: "The character of human existence itself moves us to seek explanations in that which transcends humanity. As the *Phaedo* makes clear in its focus on death and immortality, human existence is characterized by an awareness of incompleteness, an awareness that confronts us most powerfully when we face our mortality." While Stern attributes this awareness to a non-dogmatic, thereby rational choice, the present discussion attributes it to the very act of philosophizing itself leaving intact the erotic, and thereby transcendent, character of the act—an act committed not by Plato but by Nietzsche as well, his protestations and those of his epigoni notwithstanding. In attributing this awareness to "Socratic rationalism" (an infelicitous anachronism), Stern seems to regard rationalism as an accomplishment in itself. The major difference between Stern's argument and the one offered here is that he locates the "ground" of Socratic philosophy in self-understanding (p. 180). This seems inconsistent with what he claims in the above-quoted passage from p. 145. A point of departure is not a ground. Consistent with what Stern writes on p. 145, Plato discovers the ground in the *order* of the world (that order he calls divine) which is distinct from the self, but of which the self is part. The result is a firmer and more stable awareness and appreciation of politics and humanity.
2. For some sense of Plato's use of irony, see *Phaedrus* (276d), *Republic* (536c), *Statesman* (268d), *Timaeus* (59c), *Laws* (685a). Then confer with Aristotle's *Politics*, II, 6 (1265a6) and Kierkegaard's *Concept of Irony*.
3. All of the quotations from Plato's dialogues in this essay are taken from Plato, *The Collected Dialogues*, edited by Edith Hamilton and Huntington Cairns (New York: Pantheon, 1961).
4. Eva Brann, "Commencement Address," St. John's College Magazine (Annapolis: July, 1974). Cf. Aristotle's *Politics*, Bk. II, ch. 8, near the end of the chapter.
5. Julien Benda, *La trahison des clercs*, 1927; Paul Johnson, *The Intellectuals* (New York: Harper & Row, 1989).
6. Thomas Aquinas, *In Aristotelis Librum De Anima Commentarium* (Turin: Marietti, 1948), I, VIII, 107, p. 31. [Plato] "had a bad method of teaching."
7. Aristotle, *Metaphysics*, Bk. I, ch. 6, 1987b1.
8. Ann Hartle, *Death and the Disinterested Spectator* (Albany: State University of New York Press, 1986), pp. 81-83.
9. Plato, *Apology*. (32a). "The true champion of justice, if he intends to survive even for a short time must necessarily confine himself to private life and leave politics alone."
10. Plato, *Ibid.*, (29c-31b).
11. An afterlife for the *individual self* is represented in the *Phaedo* in the form of a myth (107d-114d).
12. The last lines of the *Phaedo* read as follows: "Such, Echecrates, was the end of our comrade, who was, we may fairly say, of all those whom we knew in our time, the bravest and also the wisest and most upright man" (115e). The theme of the *Phaedo* corroborates that very famous line in the *Apology* which has been incompletely and therefore incorrectly translated as "The unexamined life is not worth living. The Loeb Library edition of the Greek text says, ὁ δὲ ἐνεὶ τᾶστω βίῳ οὐ βιωτὺν ἐστὶν ἄνθρωπος (The unexamined life is not worth living for a man). (38a5-6). Plato connects inquiry necessarily with being human. For God (or the gods) and beasts philosophy is unnecessary, indeed, it is absolutely useless. It is, however, useful and necessary in order to be human. Cf. Aristotle, *Nicomachean Ethics*, Bk. X, ch. 7-8, 1177a10-1179a30. Non pace to Pragmatists, Marxists, Darwinians, Freudians and the like who either deny theoretical inquiry or reduce it to the practical or the emotive.
13. In chapter 14 of *The Birth of Tragedy*, Nietzsche suggests another interpretation of the *Phaedo*, another interpretation of Socrates' life and death and consequently another conception of human nature. Man for Nietzsche is *homo faber*, man the maker—the maker of illusions, the constructor of paradigms, the creator of fictions. Those illusions or fictions are, of course, science, religion, history, mathematics, and indeed philosophy—fictions whereby human beings keep the absurdity and horror of nature at bay. Nietzsche suggests that in the face of death, Socrates, that "despotic logician," senses the limits of reason when voices in his dreams urge him to practice and cultivate the arts. The *Phaedo* reports that Socrates believes that he was already practicing the greatest art, i.e., philosophy, but that in order to clear his conscience and to

obey the dream, he set himself the task of writing poetry by versifying some of Aesop's fables. Socrates points out that a poet who is worthy of the name ought to work on imaginative themes, not descriptive ones. He goes on to say that since *he* was not good at inventing stories, he made use of the already existing fables and versified them (60-61c). Nietzsche argues that this passage shows that both Plato and his character, Socrates, recognize that it is poetry, the creating of fiction, and not philosophy that defines human nature. Better yet, philosophy is just a sub-category of poetry with which one faces death by making existence appear intelligible and thereby justified. If, however, Nietzsche's argument is sound, then we cannot take it seriously. We must regard it as just another fiction. On the other hand, if we take his argument seriously, then we must regard his thesis, i.e., that humanity is defined by fictions, by illusions, by poetry, as unsound. [The prohibition against testing a thinker's thesis on his own thought, i.e., the prohibition against self-reference, does not *derive* (*pace* Bertrand Russell) from the intrinsic course of an argument, but from an extrinsic wish to control or protect it by an *ad hoc* regulation.] For Nietzsche's own argument is not a poem; it is a philosophic statement, false though it may be.

Again, to say that poetry defines human nature is philosophizing, not poetizing. Furthermore, poetry as Nietzsche himself claims, aims at domination of nature through the creation of fictions. The poet is a "fabricator" like the gods. Philosophy, however, seeks to *understand* nature (and everything else—including poetry) in order to say something true about them. This is not to deny that philosophy has its own rhetoric. Philosophizing, i.e., inquiring about things in order to understand them, is at the same time a recognition of the finitude, the mortality, and the humanity of the philosopher.

Finally, what Socrates does in the face of death, as the *Phaedo* bears out, is not poetry but philosophy. The myths that supplement the arguments for immortality are not higher truths but ways of showing the limits of reason and, therefore, the limits of human nature.

14. Heidegger argues this same point in his discussion with Cassirer at Davos in 1929. Cf. "A Discussion between Ernst Cassirer and Martin Heidegger," tr. Francis Slade, in *The Existentialist Tradition*, Ed. Nino Langiulli (New York: Doubleday, 1971), pp. 192-203, reprinted as *European Existentialism* with a new introduction, (New Brunswick: Transaction Books, 1997.)

# Dispensing with (Men of) Reason<sup>1</sup>

Stephen Hetherington  
The University of New South Wales

Many philosophers of feminism believe that Genevieve Lloyd, in her book *The Man of Reason*,<sup>2</sup> struck a significant blow against traditional analytic epistemology.<sup>3</sup> Supposedly, she did this by showing that the concepts of reason and rationality have, in the historical hands of Western philosophy, become *genderized*—in particular, masculinized. In this paper, I will not comment on whether other areas of philosophy are guilty of that charge; my present aim is simply to defend epistemology in particular against it.<sup>4</sup> I will argue, first of all, that Lloyd's discussion in particular—whatever its other merits might be—poses no fundamental threat to epistemology.<sup>5</sup> Although her reasoning is historicist, it fails to establish the historical links it needs to establish; indeed, it unwittingly provides some confirmation of epistemology's *enduring* importance. But I will also argue, more fundamentally and more generally, that *no* feminist argument for Lloyd's kind of conclusion could pose any real threat to epistemology: The sort of genderization which Lloyd and others believe that she has uncovered would be most clearly explicable as being a failing as such, only *via* some standard elements of epistemology—which is to say that whatever shortcomings, if any, such feminist discussions could ever uncover can easily be absorbed *within* the scope of standard epistemological explanation. Epistemology's own metaphilosophical potential will thus be seen to confirm its own continuing significance, even in the face of feminist-historicist arguments like Lloyd's.

## I

How influential has Lloyd's book been? Among those philosophers who regard what they call *feminist* epistemology as being important, the answer seems to be "Very influential indeed." Here are some representative remarks in support of that answer.

Klein, who is *not* approving of Lloyd's book, says (1996, 61) that it is the work usually cited as evidence of the historical male-bias of epistemology.

Code is one who often cites Lloyd's book in that way, indeed as showing the *current* male-bias of epistemological discussions:

Lloyd shows that ideals of Reason, throughout their shifting and evolving history, designate

what it is to be a good knower, determine what counts as knowledge and as a proper object of knowledge . . . . In short, these ideals have had a tacit yet constitutive effect on the shape of western metaphysics, epistemology . . . : an effect that has filtered through into popular conceptions of what knowledge is, who knowers are, and whose knowledge claims are authoritative. (1991, 119)

Indeed, feminists have uncovered notable coincidences [which] . . . are the twentieth-century manifestation of a persistent historical practice of defining reason, rationality and objectivity through the exclusion of attributes and traits commonly associated with femininity (cf. Lloyd, 1984). They leave no doubt that the invisible knowing subject in mainstream epistemology and philosophy of science is implicitly male. (1992, 140)

Gatens (1991, 94) is equally convinced by Lloyd:

phallogentrism operates by way of dichotomous thought, where one central term defines all others only in terms relative to itself.

A recent example of feminist critique which confirms the foregoing analysis of the way dichotomies function in the history of Western philosophy is [*The Man of Reason*].

Okruhlik and Harvey (1992, 2) refer to

some of the works that have set the agenda for present efforts to articulate the relationship between women and reason

and then tell us that "one of the most important of these is [*The Man of Reason*]."

A recent anthology on Australian feminism continues that theme. Among the contributors who refer glowingly to Lloyd's book (none have any criticism of it), Diprose (1998, 91) says that

As Genevieve Lloyd has demonstrated, the history of philosophy is governed by ideals of reason and rationality that, despite pretensions to sexual neutrality, are in effect male . . . . Lloyd

argues that reason and rationality, having no meaning or value in themselves, are defined in opposition to femininity or what is considered feminine behaviour. What emerges from this process are historically specific notions of reason that are, by definition, male.

I will not belabour the point any further. It is clear that there are philosophers to whom Lloyd's discussion *establishes*, or *demonstrates*, or *confirms*, or *reveals*, or *shows* a seemingly significant *result*—these being standard 'success words', of course, and ones that are standardly applied by these philosophers to Lloyd's discussion—about those most central of epistemological concepts—reason, rationality, knowledge.<sup>8</sup> Yet we should bear in mind that the history of philosophy has contained very few—if any—non-technical books that have achieved such success. Even a well-regarded book will be thought by at most only a few philosophers to have *shown* or *demonstrated* or *established* what it claims to have shown. Substantive philosophy is simply not as simple as that. On inductive grounds, therefore, it would be rather surprising if Lloyd's book really had established what these philosophers assume that it has established. I will argue that it has done no such thing.

## II

Lloyd's conclusion is that the concepts of reason and rationality, as used by philosophers, have become *male* concepts. What does she mean by this? What characteristics does she take to indicate such genderization of those concepts? In her view, how does such genderization occur?

She thinks that it has occurred because the philosophical study of reason and rationality has, over the ages, involved the *exclusion*, from the domain of reason and the rational, of supposedly female characteristics. This has meant the "denigration" of "the 'feminine'" (p. 107). It has occurred via philosophers having "readily identified with maleness" what they have favoured (p. 104).

As to what specific characteristics have been thus favoured, one of Lloyd's clearer examples comes from her remarks on Francis Bacon. He thought of knowledge as a matter of gaining *control* over Nature (p. 10), although not by the knower's doing so in a rough or violent way. Scientists, seeking knowledge of Nature, should treat it with "the respect appropriate to a femininity overlaid with long-standing associations with mystery" (p. 17). This "give[s] a male content to what it is to be a good knower" (*ibid.*). It is male to dominate; to know is to dominate; so, to know is male.

But Bacon's is hardly an oft-heard voice in contemporary epistemology (to put the point mildly). Metaphors of domination or power are *not* part of its attempts to understand the nature of evidence, or to answer sceptical denials of our having knowledge of a world around us, for instance. A contemporary epistemologist is more likely to conceive of knowledge as being, say, a true belief which has been *reliably* acquired (i.e., acquired in a way which is likely in general to lead to the person's having true beliefs), or as a true belief which *coheres*—harmonises—with the person's other beliefs, maybe with the beliefs of some surrounding community.

This is not to imply that the contemporary analyses are epistemologically *superior* to Bacon's (or, for that matter, to any of the others discussed in this paper). It is only to weaken whatever historicist chains would be deemed by Lloyd to *bind* the contemporary epistemologists to the various long-dead epistemologists upon whom she focuses. Lloyd seeks a *conceptual* result (and literally so, with it being a result about our some of our central epistemic concepts); and she seeks it on the basis of an *historicist* argument. But already we are finding something which we will continue to find as we scrutinise Lloyd's reasoning—which is that *even by her own historicist standards* her argument will fail to have any significant conceptual implications for contemporary epistemology. This is not because contemporary epistemology is without flaws, or because it is in all respects the best epistemology that philosophy has ever embraced; rather, it is because the alleged historical failings described by Lloyd have not in fact sufficiently shaped the conceptual content of contemporary epistemology. This is clear in the case of Bacon (and it will continue to be clear in Lloyd's other case studies, as we will see soon).

Lloyd would disagree with that. After all, she says (p. 7) that Bacon's model of knowledge is "overlaid" on Plato's, and that Plato's model "has been highly influential in the formation of our contemporary ways of thinking about knowledge." And this is significant because "Plato's picture" says (*ibid.*) that

knowledge is a contemplation of the eternal forms in abstraction from unknowable, non-rational matter. The symbolism of dominance and subordination occurs in the articulation of the process by which knowledge is gained. Knowledge itself is not seen as a domination of its objects, but as an enraptured contemplation of them.

But this, too, is *not* a model that is found explicitly in contemporary analytic epistemology,<sup>10</sup> and Lloyd gives no textual support for her claim that it "has been highly influential" in the genesis of that domain.<sup>11</sup> Even at this stage of the discussion, therefore, it is hardly obvious that Lloyd is entitled—*even by her own historicist standards*—to infer that contemporary epistemologists treat Reason as male, in the sense gestured at by her.

### III

I take it that Lloyd would see the pertinent and putative genderization as lying *beneath* the current epistemological surface. If so, though, arguments are needed on her part to make epistemologists aware of this failing of theirs. What is Lloyd's argument?<sup>12</sup>

In either explaining, or arguing for, her conclusion, Lloyd says (p. 108) that because philosophers "have been predominantly male," "the conceptualization of Reason has been done exclusively by men." And hence, she says (*ibid.*), "it is not surprising that the results should reflect their sense of Philosophy as a male activity."

Ironically, though, Lloyd herself provides us with a clue to why the latter

claim of hers is inadequately supported by her former claim. For she tells us that (*ibid.*)

despite its aspirations to timeless truth, the History of Philosophy reflects the characteristic preoccupations and self-perceptions of the kinds of people who have at any time had access to the activity.

And, we must ask, what “kinds” of people have engaged—and still engage—in philosophical reflection on reasoning? Lloyd’s answer is “Men.” But another answer—if anything, a *more* obviously true one—is “Philosophers, specifically epistemologists.”<sup>13</sup> And isn’t it quite possible that even male epistemologists *have*, as a preoccupation and self-perception when doing philosophy, a willingness to transcend their maleness? It certainly is. As we saw just now, Lloyd thinks that it is the epistemologists’ “characteristic preoccupations and self-perceptions” that matter; but, insofar as one is being epistemological, one’s self-perceptions and preoccupations should include one’s trying to *escape* one’s personal circumstances as much as is possible and reasonable, in order to reflect as fairly as possible upon as many of us as is possible and reasonable. That is part of the *method* of analytic epistemology (a method practised by many female epistemologists too, it should be borne in mind).<sup>14</sup>

We should distinguish between the epistemologist *qua* person and the person *qua* epistemologist. A male epistemologist’s failings *qua* person are not necessarily his failings *qua* epistemologist. Insofar as someone is being epistemological, he or she is theorising; even if the person is male, he is not—insofar as he is being epistemological—being male. More specifically, the (male) person’s theorising will, hopefully, be done in a way that reflects his being successfully preoccupied with living up to his perception of himself as someone who is reflecting in a way that respects and recognises all individual differences and similarities. But Lloyd is treating male epistemologists as males, rather than as epistemologists. And that is unreasonably and irrelevantly selective on her part. Unless actual epistemologists have in *practice* warranted such an otherwise irrelevant treatment, her reasoning is inapplicable to them.

#### IV

What Lloyd must do, then, is show that in practice enough philosophers have let their philosophy reflect their maleness.<sup>15</sup> And this is (almost) what she seeks to do. She tells us about the Pythagoreans, about Plato, Aristotle, Bacon, Augustine, Aquinas, Descartes, Hume, Rousseau, Kant, Hegel, Sartre and (as someone who has not escaped the supposedly unfortunate influence of such thinkers) de Beauvoir. In all of these cases, it seems, gender-biased attitudes or formulations are not hard to find, according to Lloyd.<sup>16</sup>

But of how much interest is this (even assuming, as I am doing for the sake of the immediate argument, that it is true) to a *contemporary* analytic epistemologist? It is of real interest only if the following three conditions are all satisfied: (1) Lloyd’s list includes enough actual thinkers who have sufficiently influenced analytic epistemology; (2) her discussions of those relevant

thinkers concern their *epistemology* (rather than mainly, for example, their ethics); and (3) her discussions of those thinkers' epistemologies *criticise* their epistemologies as such. If these three conditions are not jointly satisfied by Lloyd's reasoning, then it fails, even by its own *historicist* standards, to establish any significant and worrying result about contemporary epistemology (even assuming, still for argument's sake, that her reasoning establishes anything perturbing about any of the historically *prior* epistemologies upon which she fixes her critical gaze).

And (as section V will explain) these three conditions are not jointly satisfied by Lloyd's reasoning. (1) and (2) are satisfied only by her discussions of Descartes and of Hume; but those discussions do not have anything critical to say about those theorists' epistemologies as such, and hence do not satisfy (3).

Part of the problem is that Lloyd blurs the borders of these philosophers' epistemologies. I take it that this is because (p. xviii)

Reason has figured in western culture not only in the assessment of beliefs, but also in the assessment of character. It is incorporated not just into our criteria of truth, but also into our understanding of what it is to be a person at all, of the requirements that must be met to be a good person, and of the proper relations between our status as knowers and the rest of our lives.

Whether or not this is a fair complaint against the philosophers Lloyd discusses, though, it mischaracterises how most contemporary analytic epistemologists approach their tasks. Part of the aim of most such epistemologists is—for better or for worse—to *detach* from such matters as a person's character.<sup>18</sup> This is not to say that a person's character is automatically of no interest to epistemologists. Character can be epistemically significant, but what makes it epistemically significant is its relationship to some *further* characteristic—a *purely* epistemic one. That is, an aspect of one's character would be deemed by such epistemologists to be epistemically significant only insofar as it tends to further or hinder one's satisfaction of some purely *epistemic* characteristic. For example, one's honesty would be epistemically significant if it is part of why one's beliefs are more *reliably* formed, or if it is part of how one uses one's *evidence* in belief-formation. It is not the character as such that matters to epistemologists, though; it is how well a person, in virtue of having that character, instantiates overtly epistemic properties that matters. There is a link, but it is not identity, between character and epistemic properties.<sup>19</sup>

So, there is already reason to suspect that Lloyd's arguments—regardless of how effective they are against philosophers who *have* tried to intertwine their views about reason and rationality with their views about character—are inapplicable to at least most analytic epistemologists (who either do not theorise about character, and hence who seek to be detached about such matters, or who think of personal character only *qua* epistemic character). Let us examine whether this is in fact so.

## V

Lloyd will presumably deny that any epistemologists really *can* detach in the way attempted by analytic epistemologists. But consider her discussions of Descartes and of Hume. She admits, in effect, that they *were* able to detach themselves in the way attempted by most contemporary analytic epistemologists. We can see this by seeing how insubstantial are her charges against them.

Descartes's sin, it seems, is to have lived when he did. Lloyd concedes (p. 48) that Descartes's epistemological method might have been "essentially private and accessible to all," and that he "thought his account of the mind opened the way to a newly egalitarian pursuit of knowledge." But, notes Lloyd, in *fact* women did not have equal access to education and science: "impinging social realities" (p. 49) still meant inequality of intellectual opportunities, which in turn meant unequal, socially determined, capacities to *utilise* the Cartesian method. Descartes, in thinking of Reason as being a kind of pure thought, distinguished "the ultimate requirements of truth-seeking from the practical affairs of everyday life" (*ibid.*). And the problem with this, according to Lloyd, was that he therefore "reinforced already existing distinctions between male and female roles, opening the way to the idea of distinctive male and female consciousness" (pp. 49-50). Consequently, the sense in which, according to Lloyd, Descartes "reinforced" prior social disparities between "male and female roles" is one that is *extrinsic* to Descartes's own thinking—to the content of his account. The latter was gender-neutral, but society was not.

Apparently, Hume is similarly 'guilty.' The most that Lloyd can say against him is that (p. 56)<sup>20</sup>

in its social context. Hume's version of Reason, like Descartes's, which made it possible, takes on associations with maleness, even if these are not specifically required by their philosophical theory.

But these are certainly not good reasons for criticising Descartes and Hume as *philosophers*. Lloyd is admitting that their abstract reasoning—its content—is itself innocent of maleness; she is claiming only that their reasoning could be *used* in what she would call a Male way (so as to locate Reason in the world in such a way as to think of females as being less possessed of reason). She is not even saying that Descartes and Hume themselves, as philosophers, ever used their reasoning in that way; she is saying only that others might well have done so.

Yet can't we turn her reasoning on its head, and urge that her own observations provide good reason for *approving* of Descartes and Hume as philosophers? In other words (and once more), isn't her reasoning failing even its *own* supposed standards? Presumably Lloyd is just assuming that being detached is—even if achievable—not desirable in a philosopher, for to make that assumption would be badly question-begging. And her description of Descartes and Hume as in effect offering views which (i) are in themselves reasonable, but (ii) are used in a biased way by others who are thinking less detachedly, is an argument *for* trying to reason in that detached a manner! If, say, Descartes's social context—comprising non-philosophers (plus philosophers who are not

being philosophical in Descartes's way)—is faulty as regards the opportunities available to, and the attitudes taken towards, women, then surely one reasonable aim a theorist might have is to detach from such contexts.<sup>21</sup> One would try to describe reason in ways that do not reflect or endorse such regrettable social realities. And Lloyd seems to admit that Descartes and Hume are not only trying to do this, but are *succeeding* in doing so. For she admits that their theories are *in themselves* neutral regarding women. It is not reasonable, therefore, for her to conclude that their *concepts* of reason—let alone ours (simply because theirs!)—are Male.<sup>22</sup>

The most that follows from her argument is that *non*-philosophers (plus non-Cartesian and non-Humean philosophers) were Male-biased, since these were the people supplying the unfortunate “impinging social realities.” But again, this is, if anything, an argument for doing philosophy—and *for* doing it along Cartesian or Humean lines (as, very broadly speaking, many analytic epistemologists already do)! I take it that a quite reasonable philosophical account can fall upon socially deafened ears, without the philosophical account thereby being a party to, or even being influenced by, the forces of such deafness.<sup>23</sup>

No doubt inadvertently, Lloyd is revealing part of why it is so important to do philosophy—and to do it in the detached way which analytic epistemologists aim to manifest. It is important because if people *could* detach from (i) (gender-)unreasonable social surroundings, to (ii) (gender-)reasonable conceptualisations, they would be exhibiting a kind of escape from such social failings.<sup>24</sup> Not only that, but they might even be able to rectify those social failings. They could do so, in the only way that detached philosophers can ever reasonably hope to change such matters. Specifically, they could *teach*—exemplify, display, impart—more reasonable ways of reasoning. For instance, they could teach a gender-neutral cognitive method such as Descartes's. (And they could inquire into its failings—but they could do so *in* a gender-neutral way. This is part of what analytic epistemologists already strive to do—and generally succeed in doing.) If more of Descartes's contemporaries, say, had applied his method, then—*ipso facto*—the social realities referred to by Lloyd would not have been so uncongenial to, or unchanged by, his neutral, abstract conception of reason.<sup>25</sup>

## VI

Lloyd implicitly concedes this. She says (p. 109) that

philosophers can take seriously feminist dissatisfaction with the maleness of Reason without repudiating either Reason or Philosophy. Such criticisms of ideals of Reason can in fact be seen as continuous with a very old strand in the western philosophical tradition . . . . Philosophy has defined ideals of Reason through exclusion of the feminine. But it also contains within it the resources for critical reflection on those ideals and on its own aspirations.

Thus, Lloyd seems to take her reasoning—her “critical reflection”—to be philosophical and, presumably, rational.

But won't *it* therefore be Male too (according to Lloyd)? That is, according to what it says itself, won't it—in particular, the concepts it employs—have been defined or constructed over time in a way that respects maleness to the detriment of “the feminine”? And (harking back to section III) philosophy is still practised by more males than females. So, wouldn't even Lloyd's own philosophical thinking (in her view) be unable to do anything *but* manifest such limitations?

The only way it can escape this limitation is if individual philosophers (*qua* philosophers), such as Lloyd herself, can do so—and if this can somehow constitute philosophy's doing so. But to admit this possibility is (i) to agree with my rejection, in section III, of Lloyd's *generic* dismissal of traditional, male-dominated, philosophy, and (ii) to concede that—contrary to Lloyd's longer, ‘case-studies argument’—individual philosophers can rise above their surrounding social realities, and still be doing (good) philosophy. In short, to make that admission is to reject Lloyd's explicit arguments. The fact that she implicitly makes it herself reveals her overall position to be inconsistent.

## VII

Lloyd's argument fails in that way, in part because *no* argument like hers can achieve its ultimate aim. Let us now see why that is so.

Purely for the sake of argument, suppose that some feminist argument like Lloyd's *could* show that reason and rationality are genderized concepts. What kind of *failing* would thereby have been established, regarding epistemology and its concepts of reason and rationality? Lloyd (along with many of the other feminist philosophers cited earlier) seems to assume that for a concept to be genderized is for it to be *unfairly* genderized, and hence unfairly used—in brief, that the concept is somehow inadequate. But in what *epistemological* way, at least, would it be inadequate?<sup>26</sup>

Is it, for example, that when epistemologists use a genderized concept of reason to make attributions or denials of reason, their claims are all false? When employing a genderized concept of rationality to describe one person as rational, and another as irrational, is an epistemologist automatically saying something *false*? I do not think so. If, for example, the concept of physical strength had been developed in a way that was thought to recognise only male physical capacities, this would not entail that nothing is strong. Nor must the genderized concept somehow be meaningless, so that all references to reason are *meaningless*. We could still understand uses of a genderized concept of reason, even if we disagreed with all of the uses (regarding them as false). And not even all uses of the genderized concept in relation to the ‘*excluded*’ gender must be false or meaningless: Even if the concept of physical strength had been developed in a way that was thought to recognise only male physical capacities, this would not entail that no females are strong.

In fact, it could happen to be the case that a specific concept which has been genderized in its history is *just as* accurately applied to members of the excluded gender as to members of the non-excluded one. Why is that? First, I see no evidence (and Lloyd provides none) that it is not the case that as many

women as men should be classed as being rational, relative to current epistemological discussions of rationality, for instance. Second (and more importantly), even a genderized concept *could* be the correct one to use in general. To categorise it as gendered is to comment on its genesis (notice Lloyd's historicism, hence her concentration on the *history* of philosophical talk of reason)—yet even a concept acquired by a selective or socially prejudicial method might turn out to be the correct concept. That is, it could turn out—in spite of its somewhat unfortunate history—to be the one which someone who wishes to make true claims about rationality should use. If this does not seem clear, consider a simple analogy: As epistemologists routinely point out, a person can have a belief which happens to be *true*, in spite of acquiring it via a quite unreliable method, such as guessing. Genesis is one thing, truth another.

Still, that is not to say that there would be nothing wrong with using a concept that possesses so unimpressive a pedigree. There is—and standard analytic epistemology *itself* allows us to understand (as follows) what that failing would be.

If Lloyd were correct in her main claim that the concepts of reason and rationality had been genderized by their development over the years, this would imply that they had been gained or formed in such a way as to reflect only a *limited perspective's* attempts at understanding or inquiry. (In Lloyd's view, those concepts would have been developed via a process which reflected male concerns more than female ones.) And the most clear-cut problem with this is that it affects, not necessarily the *truth* of the claims we make using those concepts, but our epistemic *justification* for thinking that we are making true claims using them. If our concepts of reason and rationality had been gained in so limited a way, it would be reasonable to say that attributions made using them are epistemically *unjustified*. The attributions might happen to be true, but this would not render them justified. Succeeding in using them to capture truths about people's cognitive capabilities would be like reaching a true belief by guessing. And analytic epistemologists routinely tell us that one way to lack justification for one's beliefs is to acquire them, or the crucial concepts contained in them, by employing perspectives that are too restricted in scope to reflect enough of those aspects of reality that could be relevant to determining the truth-value of those beliefs.

There is a profound irony in this. Many analytic epistemologists (certainly internalist ones)<sup>27</sup> would describe a lack of justification as a lack of rationality—a lack of theoretical rationality.<sup>28</sup> They would therefore understand the use of genderized concepts as being a failing—but they would do so in terms of their concept of rationality. And this shows that the failing—even if it exists at all—could never be one that tears at the heart of epistemology. *It is an epistemic failing itself—an eminently epistemologically explicable one.* Any proved genderization of a concept of rationality—or, more plausibly, of some specific conception(s) of rationality—could easily be accepted by epistemologists as being a failing; yet it would be so easy for them to agree that it is a failing, precisely because they can *explicate* the failing *qua* failing, by calling on the very concepts of reason and rationality which feminists like Lloyd think that they are undermining by uncovering the supposed genderization. Thus, even *if* epistemologists were to accept that a good argument could be given for the con-

cepts of reason and rationality having been genderized, they would not thereby be obliged to discard those concepts.

Accordingly, the most that a feminist could hope to show by arguing for Lloyd's sort of conclusion is that *one more way* for claims about reason or rationality to be unjustified—irrational, even—is for them to be genderized. An ineliminable problem for feminist projects like Lloyd's is that any such result is perfectly understandable as a detail *within* standard epistemology—a detail explicable *via* the concepts of rationality and of justification, rather than one that throws into doubt the value or coherence of those concepts. It certainly does not entail that no claims about reason or rationality are justified or rational, let alone true. Once more, therefore, even at best the feminist arguments give us simply some reason to absorb their concerns about reason and rationality *into* mainstream epistemology. Even *at best* we are being given just some more *examples* of considerations to bear in mind when assessing a belief's justifiedness, say, rather than some fundamental challenge to the very concept of justifiedness. We can sensibly ask this: "Has that belief been reliably acquired—by, in part, the believer's having been sensitive to *different* perspectives, including male and female ones?" We cannot sensibly say this: "Because genderization has occurred in how some epistemologists have thought about reason and rationality, no belief is rationally held—and, indeed, the concepts of reason and of rationality are themselves epistemologically empty or inadequate concepts." So, feminist worries like Lloyd's pose no threat as such to mainstream epistemology: *Even if their accusations of genderization are ever true, this never entails that something fundamental about epistemology itself needs to be changed.* At most, therefore, these feminist worries could direct us towards an interesting but comparatively small area to which we may apply some broader and more fundamental concepts from within standard epistemology. That is, at most the feminist arguments might help us to describe some useful but comparatively restricted details within the explanatory scope of standard epistemology.

Antony (1993, 187), for one, would be sympathetic to that conclusion:

If we focus on the existence of what might be called a "feminist agenda" in epistemology—that is, if the question, "Do we need a feminist epistemology?" is taken to mean, "Are there specific questions or problems that arise as a result of feminist analysis, awareness, or experience that any adequate epistemology must accommodate?"—then I think the answer is clearly yes. But if, taking for granted the existence of such an agenda, the question is taken to be, "Do we need, in order to accommodate these questions, insights, and projects, a specifically feminist alternative to currently available epistemological frameworks?", then the answer, to my mind, is no.

## VIII

At least in principle, then, epistemologists need not resist feminist attempts to inquire into whether there has been implicit genderization in various uses—even epistemological uses—of the concepts of reason and rationality. Crucially, though, epistemologists need never *fear* the results of such inquiries. That is because no such discovery of some such genderization could ever reveal the *inadequacy*—not, at least, for epistemological purposes—of the concepts of reason and of rationality. The most that could be revealed is that some people—perhaps even some epistemologists—have lacked some justification which they might have believed they possessed. But this can be understood as being a *failing*, only relative *to* a coherent and stable epistemological concept of justification: To uncover and to understand such failings is part of the standard epistemological *use* of that concept. An epistemologist would use a concept of rationality, say, in order to understand why any putative genderization is at all a failing. Consequently, the feminist would have provided no reason to *discard* the concept due to the existence of the failing. What arguments such as Lloyd's reveal—inadvertently, of course—is that what is needed, instead, is better and more widespread epistemological training in how to use that concept, rather than a dismissal of the concept. Society requires *more*—not less—epistemology. More people need to understand better the breadth and depth of standard epistemology.

Thus, an epistemologist, by using the concepts of reason and rationality, can agree that the unfortunate moves in the name of reason which Lloyd claims to have revealed would indeed be failings.<sup>29</sup> The epistemologist would thereby view Lloyd's putative examples of gendered uses of the concepts of reason and rationality as being examples of how to *misapply* the concepts. However, he or she need not therefore *jettison* the concept. On the contrary: By *retaining* that concept, he or she can give the feminist data whatever explanatory welcome they would deserve in this setting; in that way, they may be understood to *be* data that indicate the existence of some sort of a failing. As section VII explained, epistemologists themselves—by *using* the concepts of reason and of rationality—can understand perfectly comfortably just what the failing would be in genderizing the concepts. (The failing would be that epistemic justification is absent.)

Lloyd herself would probably not interpret her argument in that way.<sup>30</sup> If so, though, she still owes us an account of just what epistemological *failing* she has described (and claimed to have uncovered). And the same is true of those many other feminist philosophers who speak in terms similar to Lloyd's. Still, perhaps it is not surprising that they have supplied so little analysis of what epistemological failing would be present and of why, according to them, reason and rationality are concepts whose life has been lived in vain. After all, if I am right, no such analysis could be correct: No epistemologist needs to regard any argument like Lloyd's as posing a fundamental or underlying threat to his or her discipline. In that regard, at least, Lloyd's argument fails, as do all other feminist arguments of similar intent and support. Regardless of whether or not some such arguments should worry other philosophers, epistemologists may view them with equanimity.

**REFERENCES**

- Alcoff, L. M. (ed) 1998. *Epistemology: The Big Questions*. Blackwell.
- Antony, L. 1993. "Quine as Feminist: The Radical Import of Naturalized Epistemology." In *A Mind of One's Own*, (eds) L. M. Antony and C. Witt, at 185-225. Westview Press.
- Atherton, M. 1993. "Cartesian Reason and Gendered Reason." In *A Mind of One's Own*, (eds) L. M. Antony and C. Witt, at 19-34. Westview Press.
- Braidotti, R. 1991. *Patterns of Dissonance*. Polity Press.
- Code, L. 1991. *What Can She Know?* Cornell University Press.  
1992. "Feminist Epistemology." In *A Companion to Epistemology*, (eds) J. Dancy and E. Sosa, at 138-142. Blackwell.
- Coombs, A. 1996. *Sex and Anarchy*. Penguin.
- Diprose, R. 1998. "Ethics." In *Australian Feminism*, (eds.) B. Caine (general), M. Gatens, E. Grahame, J. Larbalestier, S. Watson, and E. Webby, at 90-98. Oxford University Press.
- Gatens, M. 1991. *Feminism and Philosophy*. Polity Press. 1998. "Sociability and Inclusion." In *Australian Feminism*, (eds.) B. Caine (general), M. Gatens, E. Grahame, J. Larbalestier, S. Watson, and E. Webby, at 295-303. Oxford University Press.
- Goldman, A. I. 1986. *Epistemology and Cognition*. Harvard University Press.
- Grosz, E. 1990. "Philosophy." In *Feminist Knowledge*, (ed) S. Gunew, at 147-74. Routledge. 1992. "What is Feminist Theory?" In *Knowing Women*, (eds) H. Crowley and S. Himmelweit, at 355-69. Polity Press.
- Gunew, S. 1990. "Feminist Knowledge: Critique and Construct." In *Feminist Knowledge*, (ed) S. Gunew, at 13-35. Routledge.
- Haack, S. 1993. "Knowledge and Propaganda: Reflections of an Old Feminist." *Partisan Review* 60: 556-64.
- Held, V. 1993. *Feminist Morality*. University of Chicago Press.
- Hetherington, S. C. 1996. *Knowledge Puzzles*. Westview Press.
- Klein, E. R. 1996. *Feminism Under Fire*. Prometheus Books.
- Koertge, N. 1996. "Feminist Epistemology: Stalking an Un-Dead Horse." In *The Flight From Science and Reason*, (eds) P. R. Gross and N. Levitt, at 413-19. New York Academy of Sciences.

Lehrer, K. 1990. *Theory of Knowledge*. Westview Press.

Lloyd, G. 1979. "The Man of Reason." *Metaphilosophy* 10: 18-37. Reprinted in *Women, Knowledge, and Reality*, (eds) A. Garry and M. Pearsall (1989), at 111-28. Unwin Hyman. 1984. *The Man of Reason* (first edn). Methuen. 1993a. *The Man of Reason* (second edn). Routledge. 1993b. "Maleness, Metaphor, and the 'Crisis' of Reason." In *A Mind of One's Own*, (eds) L. M. Antony and C. Witt, at 69-83. Westview Press.

Moser, P. K. 1985. *Empirical Justification*. Reidel.

Okruhlik, K. and Harvey, E. D. 1992. "Introduction." In *Women and Reason*, (eds)

K. Okruhlik and E. D. Harvey, at 1-18. University of Michigan Press.

Pinnick, C. L. 1994. "Feminist Epistemology: Implications for Philosophy of Science." *Philosophy of Science* 61: 646-57.

Richards, J. R. 1996. "Why Feminist Epistemology Isn't." In *The Flight From Science and Reason*, (eds) P. R. Gross and N. Levitt, at 385-412. New York Academy of Sciences.

Sargisson, L. 1996. *Contemporary Feminist Utopianism*. Routledge.

Segal, L. 1987. *Is The Future Female?* Virago.

Zagzebski, L. T. 1996. *Virtues of the Mind*. Cambridge University Press.

## NOTES:

1. Stephen Cohen made some very helpful comments on an earlier version of this paper, as did a referee for this journal.
2. The book originally appeared in 1984, parts of it developing an earlier paper (1979). A second edition has since appeared (1993a), differing from the first edition in adding a new preface and a new bibliographical essay. All of my otherwise unattributed page references in this paper are to the book's second edition.
3. Little wonder, then, that Grosz (1990, 162) regards Lloyd—approvingly—as being not simply a feminist, but a *radical* feminist, in this setting.
4. Lloyd does not describe her argument as being directed against epistemology in particular. Several areas of philosophy employ those concepts, and she does not distinguish among those areas, telling us (p. 108) more sweepingly that “Philosophy” as a whole has been called into question.
5. Klein (1996, 61-63), too, has argued for this, and I endorse her criticisms. They are rather brief, though. Lloyd's account deserves a fuller treatment.
6. This does not include all feminist philosophers, of course. For some dismissals of the idea of feminist epistemology, see Haack (1993), Klein (1996, chs. 2, 3), Koertge (1996), Pinnick (1994), and Richards (1996).
7. Unusually for a contemporary philosophy book first published in English, *The Man of Reason* has been translated into other languages—the first edition into German, the second into Norwegian and Turkish. I take this to be a further mark of its influence. Part of it, too, has been reprinted, in Alcoff (1998, 387-91).
8. For some further recent examples of such approval, see Alcoff (1998, 385), Braidotti (1991, 185-90, 200), Gatens (1998, 297), Grosz (1992, 367), Gunew (1990, 13, 16, 17), Held (1993, 43-44), Sargisson (1996, 134-36), and Segal (1987, 33). The approval has even occurred outside of Academe. Coombs (1996, 80) has this to say:
 

In her classic book *The Man of Reason* [Lloyd] shows how philosophers through the ages have consistently described the duality of man/woman in . . . terms that exclude women from the world of reason.
9. On reliabilism and coherentism, see, for instance (and respectively), Goldman (1986) and Lehrer (1990).
10. Where, in contemporary epistemological accounts of the genesis or the nature of knowledge, is anything like the concept of “enraptured contemplation” to be found? Nowhere. Does the concept of dominance appear in analytic epistemological discussions of a person's using sensory impressions to gain knowledge of the presence of another person? It does not. Again, this does not entail that Bacon's conception of knowledge, say, is conceptually *inferior* to contemporary analyses. But it does weaken any *historicist* case which Lloyd could make, on the basis of supposed shortcomings in Bacon's conception, regarding claimed weaknesses of our contemporary epistemological conceptions of rationality.
11. Epistemology *is* presumably the domain about which Lloyd is talking at this point. As note 4 observed, she generally talks of philosophy in general, rather than of epistemology in particular. But it is natural to see some specific parts of philosophy—such as the part to which other parts typically delegate the job of thinking about knowledge—as being among the intended objects of at least many of her comments. And, in speaking of “our contemporary ways of thinking about knowledge,” Lloyd apparently does mean to refer to contemporary epistemology and to those who are influenced by it.
12. The bulk of it is in her chapters 1 through 6. In chapter 7, she sums up the argument. I will begin where she ends, with her remarks in chapter 7. In section IV, I turn to the rest of her argument.
13. There are philosophers whose interest in reason has not been epistemological. As we will see, these are the philosophers upon whom Lloyd tends to concentrate. As we will also see, this at worst biases—and at best dilutes—her argument against epistemology.
14. Could Lloyd say that, necessarily, such attempts will *fail*—because male epistemologists cannot overcome their maleness, even when they have learnt, and are employing, analytic methods? (Stephen Cohen asked me this question.) Clearly, Lloyd can say it. But does she provide good evidence for the existence of such a limitation? She would take herself to have done so, since she does provide examples of what she thinks are relevant failures by male philosophers. I consider that *evidence* in sections IV and V.
15. Or (cf. note 14 above) she must show that in principle this would happen, were most philosophers male. But how would one argue for that subjunctive claim? It is far from clear how to do so. The simplest approach is to do what Lloyd attempts to do: (i) she notes that most philosophers *have* been male, and then (ii) she tries to provide (enough) *actual* examples of male philosophers failing to transcend their maleness. It is this approach of hers that I am about to consider.
16. We have already noted two of those formulations—Plato's and Bacon's.
17. It is not good enough to say, vaguely, as Lloyd does (1979, 18), that “Past philosophical reflection has after all helped form our present thought structures.”
18. And isn't one of the possible morals we can learn from studying Hume, say, the *disjointness* of our use—let alone our study—of Reason from “the rest of our lives”?
19. For a recent, and more extended discussion of this issue, see Zagzebski (1996).
20. So, it is at best misleading for Lloyd to say (pp. 49-50) that Descartes's account “reinforced” prior social disparities. It coexisted with them, and (in spite of Descartes's own efforts) failed to change them. But that is hardly the same as reinforcing them. In the preface to her book's second edition, Lloyd herself seems to question this move of hers. In the first edition, she now says (p. xiv),

I put more emphasis in this section [on Descartes] than I now think appropriate on the interaction between philosophical content and

its extraneous social context. I would now emphasise, rather, the interaction between philosophical content and the implicit play of symbols which links Descartes's texts with others in the philosophical tradition.

However, Lloyd does not proceed to tell us how her argument will allow her to do this in a way that shows Descartes's discussion of Reason to be conceptually or symbolically genderized—or, a fortiori, how she will reveal a conceptual or symbolic link between the content of Descartes's discussion and contemporary epistemological discussions. (Instead, I am about to argue that her own reasoning provides a way to understand Descartes's account as non-genderized.)

21. Another reasonable aim would be to *change* the context—that is, to stay in it, while nevertheless altering its objectionable aspects. As I will explain briefly, even the detached theorist has a way of trying to achieve this aim.
22. As we saw, Lloyd makes that inferential leap by treating the theories—hence the concepts—of Descartes and of Hume as being determined largely by their failure to change prior social disparities. However, that is anything but an obviously true principle about concept content. Can't an individual use a non-standard concept? Maybe only a special individual can do so. Yet weren't Descartes and Hume special in their use of concepts? Isn't that why contemporary epistemology does, and should, respect them so much?
23. I have encountered the objection that Lloyd's admitting that Descartes's and Hume's accounts in themselves avoid genderization is just evidence of her *evenhandedness*—and, consequently, a virtue of her account. I cannot see how that claim helps Lloyd in this setting. Evenhanded in intent or not, her concession is simply inconsistent with her conclusion. After all, her conclusion is that our concept of reason has been progressively genderized by various pertinent philosophers. This conclusion requires Lloyd to trace a pattern of pertinent philosophers' accounts that have not transcended genderization. (And if the conclusion is to be applied warrantably to analytic epistemologists, for example, *those* accounts need to be ones that have influenced such epistemologists' understanding of reason.) So, for Lloyd to admit that Descartes's and Hume's accounts do, strictly speaking, transcend genderization is for her to undermine her own argument. (And the fact that it is those accounts—Descartes's and Hume's—upon which contemporary analytic epistemologists draw so heavily is, again, a strength of such epistemology—even by Lloyd's own reasoning, ironically!)
24. Witness one of the oft-claimed benefits of pertinent people functioning as good role models for others. A detached thinker—considering many possibilities and counterpossibilities, different viewpoints, competing viewpoints—might be a good intellectual role model (especially in social circumstances like, apparently, Descartes's or Hume's).
25. For more historically detailed criticism of Lloyd's discussion of Descartes, see Atherton (1993).
26. Elsewhere (1993b), Lloyd says that what she has shown (in 1993a) is that reason has been *symbolised* as male. She does not say what failings—in particular, what epistemological failings—it therefore has.
27. Hetherington (1996, chs. 14, 15) provides a brief introduction to the epistemological distinction between epistemic internalism and epistemic externalism.
28. For more on the links between theoretical rationality and epistemic justification, see Moser (1985, ch. 6).
29. Please bear in mind, though, that sections II through VI have argued that epistemologists need *not* accept that Lloyd has succeeded in revealing any such unfortunate moves having been made—not, at any rate, any that have in fact sufficiently shaped epistemology's current logical space.
30. But, of course, that does not make these remarks irrelevant. Cf. Lloyd (p. 110): "I have often highlighted points which were not salient in the philosophers' own perceptions of what they were about."

# Ethics and the Wittgensteinian Approach

Eyal Moses

In his paper “Ethics without Principles,”<sup>1</sup> Avrum Stroll argues against “a certain model about the nature of moral reasoning”: a model which “is a form of classical foundationalism; it holds that all moral reasoning rests upon a foundational principle that is the basis of other moral principles or injunctions, and that is itself not supported by evidence, reasons or any other moral principles” (Stroll, p. 310). Examples of ethical theories conforming to this model are Utilitarianism (in which the Principle of Utility serves as the foundational principle) and Kantianism (in which the Categorical Imperative serves as the foundational principle).

In contrast to this model, Stroll advocates an approach to ethics inspired by Wittgenstein's view of language in *Philosophical Investigations*. He advocates the view that “the moral world is infinitely complex and accordingly . . . that no single theory will accommodate its variety of practices and the cases falling under them” (p. 315); that cases calling for moral reasoning can't be classified in precisely definable categories based on a common feature and are instead grouped together by family resemblances; and therefore that moral reasoning consists of examining particular cases and the many factors applying to them, with no reference to fundamental principles.

In the following I will examine Stroll's arguments against the foundationalist model of ethics. I will argue that Stroll's arguments do work against the foundationalist model as he presents it, and against those ethical theories (such as Kantianism or Utilitarianism) that conform to such a model. However, Stroll has failed to refute moral foundationalism, or to establish his own “ethics without principles” as the only viable alternative, because of a basic weakness in his argument—his failure to distinguish *rules* from *principles*. Stroll's own proposed approach to moral reasoning shares the essential feature of the model he is arguing against—morality as consisting of a set of rules; as a result, Stroll's approach is itself vulnerable to Stroll's own arguments. In contrast, the Aristotelian or Objectivist approach to ethics presents a version of foundationalism which, unlike the Kantian or Utilitarian version, leads to a morality based on principles rather than rules; and consequently, Stroll's objections do not apply to it.

## What is the foundation of morality?

Before turning to Stroll's central argument, let me address a side issue that he raises in his paper against foundationalism. Stroll argues that moral foundationalism suffers from “a patent inconsistency”:

Moral reasoning is, according to the model, deeply committed to the exercise of justification. . . . Yet the principle which is ultimately appealed to in this process is stipulated to be beyond justification. . . . There is thus a kind of pragmatic inconsistency in attempting a justification while employing a principle which itself cannot be justified. On this ground alone the model can be rejected. (p. 315)

This may be a valid argument against Kantianism or Utilitarianism; but it does not work against Aristotle's approach to ethics, or Ayn Rand's, or that of neo-Aristotelian philosophers such as Machan or Den Uyl and Rasmussen; and these are also forms of foundationalism.

In the Aristotelian or Objectivist approach to ethics, evaluative propositions are derived from facts about the nature of man, or of any other living organism. These approaches do not accept Hume's "is-ought" dichotomy, which Stroll takes for granted; they hold that what any organism ought to do is determined by what it is.

On the Aristotelian view, moral judgments identify the constituents of man's flourishing. These identifications are based on observable facts of man's nature.

On the Objectivist view, moral judgments identify how man should act so as to make his own long-term survival most likely. Again, this identification is based on the facts of man's nature and what will in fact help him to survive or hinder his survival. (Survival as the standard is not itself any sort of "moral axiom which is itself beyond justification"; there is no moral principle on this view which states that "man should choose to live". Rather, ethics only applies to those people who do choose to live. Hypothetical people who don't choose to live have no reason to take any action at all, and moral judgments do not apply to them; nature will simply take its course.)

While there is a long tradition that denies the possibility of deriving evaluative conclusions from facts, Stroll is not justified, when criticizing foundationalism, in taking for granted that foundationalists accept this dichotomy. On both the Aristotelian and Objectivist views, moral conclusions are no different in principle from medical ones. They are derived from general conclusions about man's nature and the requirements of his life. These conclusions, like all factual propositions, are derived from perceptual observation and logical reasoning; and they in turn lead to principles about how man should live, which serve as the basis for specific moral (or medical) decisions. This is a form of moral foundationalism; moral reasoning on specific cases is based on general principles, and these principles are justified on the basis of more fundamental propositions. The foundation, however, is in observed facts; there is no need for a moral axiom that is itself beyond justification. Stroll's argument, therefore, fails against this form of moral foundationalism.

### **Rules vs. Principles**

Stroll uses the terms "moral rules" and "moral principles" interchangeably in his paper. However, these are two very different types of moral propositions, representing very different approaches to ethics. Stroll's failure to distinguish between them causes a major weakness in his argument. (My discussion below of the distinction is based on David Kelley, "Rules—or Principles?" *IOS Journal*,

vol. 6 no. 6, February 1997.)

A principle is “a fundamental, primary, or general truth, on which other truths depend”. (Ayn Rand, “Anatomy of Compromise”, in *Capitalism: The Unknown Ideal*, p. 144). Moral principles are fundamental, primary, or general truths about how man needs to act in order to live successfully, on which other moral truths depend. Rules, in contrast, are “self-contained prescriptions about concrete actions or situations, telling you what to do or how to do it”. (David Kelley, “Ruled—or Principled?” p. 1).

Rules are limited in scope, prescribing or forbidding a particular type of action in a particular type of situation. While they are not strictly “concrete”—i.e. they don't refer to one action on one occasion—they are very narrow abstractions, associating occasions and actions by specific aspects that are either directly perceivable or require only very simple concepts to recognize. Principles, in contrast, are broad abstractions covering a multitude of different actions and occasions, providing comprehensive guidance for a wider range of situations than any set of rules could possibly cover. “Do not smoke in elevators,” “do not commit adultery,” or “honor thy father and thy mother.” are examples of rules; “treat other people with courtesy,” “try to build relationships with others that are based on mutual trust,” or “treat people as they deserve in light of the benefit or harm they have done you,” are examples of principles.

Rules are supposed to be applied to the cases they cover with no need for exercising judgment. Principles, in contrast, because of their abstractness, have to be applied to any particular case by the exercise of judgment, taking account of the specific facts of the context. Rules, if they are not anchored in principles, cannot be rationally defended. Rule-based moralities are either based explicitly on authority rather than reason—e.g., the Ten Commandments in the ethics of Christianity—or are derived from some moral axiom which must be accepted without justification—e.g. most practical attempts to practice Kantianism or Utilitarianism, which end up as a set of rules. Principles, in contrast, can be justified in the Aristotelian or Objectivist approach to ethics, rationally defended by identifying man's needs and capacities, deriving from them man's basic requirements for living successfully.

It is common to refer to a person as “principled” if he consistently sticks in his actions to certain moral rules. There is a huge difference, however, between someone who accepts specific rules of action and then follows them regardless of the context, and someone who formulates broad principles to guide his actions and then makes concrete decisions by carefully considering the applications of his principles to the case. It is the second type of person that is more deserving of being called “principled.” Such a principled person will usually have—at least for the easier, more commonplace concrete decisions he has to make—specific rules of action derived from his principles, which he need not think about in every case, and which he will experience as ingrained habits and as emotional reactions; but he is always committed to understanding the justification of such specific habits, and to acting against them if the deeper principle and the specific context make them inappropriate.

There is one other crucial difference between the function of rules and principles, which is not illustrated by the discussion below, but which should

be noted to fully appreciate the distinction. Moral rules have only one function: to limit man's actions, identifying certain actions as wrong or as mandatory. Moral principles have a much wider function—to guide man through the fundamental choices of his life, in how to live, how to choose his purposes, and how to find happiness. Identifying actions as wrong is only a small part of that function. It is, however, the part I focus on in the discussion below, because all of Stroll's examples in his paper involve identifying actions as wrong. In order to make my answer to Stroll as clear as possible, I have chosen to illustrate how a principled approach would address the same examples that he raises.

### **Principles and Stroll's argument**

Stroll attributes several assumptions to foundationalism. The first two are the most important ones for the purposes of Stroll's objections:

First, adherents to the model presuppose that it is possible accurately to identify those cases or situations to which moral reasoning applies. It is assumed that there is something special about these cases and the reasoning that concerns them. In particular it is assumed that such cases have sharp boundaries and this is why they allow for easy identification. Truth telling or adultery or abortion are such cases. One either tells the truth or does not, one either commits adultery or does not. What counts as adultery, for example, is a matter of definition, and particular actions can be sharply characterized via such definitions. If a married person has sexual intercourse with someone other than the person he or she is married to this is defined as adultery.

Second, because such cases have sharp definitional boundaries, the model of moral reasoning contends that via the application of a fundamental moral principle—such as the Categorical Imperative—one can deduce definitive moral judgments about them. Thus, any case of adultery is always wrong, and adulterers are to be censored for their behaviour. (Stroll, p. 311)

Against these assumptions, Stroll cites the Wittgensteinian view that concepts do not identify any single quality or feature that all the concretes it subsumes have in common, and are based instead on “family resemblance”.

Even where a term like “adultery” can be applied to various cases of human conduct there is no single feature which all these cases exhibit—some common quality, or essence, for example. . . . Because there may be no such feature, we cannot do what the model implies we can always do, namely make infallible assessments about the rightness or wrongness of an act per se or about the conduct of those who participate in it. That is, I deny that from the fact that a case can correctly be described as adulterous it follows that it is necessarily wrong, or that those engaged in such an action have acted wrongly. (Stroll, p. 312)

On the purely epistemological issue, Stroll's Wittgensteinian claim, as applied to the case of the term “adultery,” is false if taken literally. Stroll says that “there is no single feature which all [cases of adultery] exhibit,” but clearly there are at least three such common features—the fact that sexual intercourse has occurred, the fact that the parties are not married to each other, and the fact that at least one of the parties is married.

At the same time, many readers would feel a strong intuition that Stroll's argument seems plausible; I myself did, when first reading his paper. What is it that makes the argument seem plausible? To understand that, let us look at

Stroll's example, adultery, and consider the difference between analyzing cases of adultery in terms of principles or of rules.

A principled approach requires identifying a principle which is relevant to cases of adultery, and which is justified by factual identification of man's basic requirements for living successfully. Such a relevant principle is: "one should seek to build relationships with others that are based on mutual trust." This is a broad abstraction, providing comprehensive guidance in conducting one's relationships with others. It also identifies a wide range of actions that are wrong, in many different situations: dishonest business dealings, deception among friends, as well as many cases of adultery. Adulterers, in most cases, try to hide the adultery from their spouses. Also, marriages usually involve a mutually understood expectation that the sexual relationship between the two spouses will be exclusive, and adultery violates that expectation. For both these reasons, adulterers are harming their ability to build a relationship of mutual trust with their spouse; applying the general principle, we can justify the judgment in any particular case that the adultery is morally wrong.

By using this principled approach, we identify what is essential to morally judging any specific case of adultery, and to guiding one's own decision if one is faced with an opportunity to commit adultery. The principle points to facts that need to be considered about the context of any specific case, which are morally more essential than "does it fall under the definition of adultery"? For example, if a married man has a sexual affair with a woman other than his wife, but does so with his wife's prior knowledge and consent, the principle helps us identify an essential difference between this case and the more common cases of adultery; while the action does fall under the definition of adultery, the principle of mutual trust does not justify judging it as wrong; anyone who claims that it is still wrong would have to defend that claim by formulating and justifying some other relevant principle. In contrast, if a couple form a long-term romantic relationship, with a mutually understood expectation of sexual exclusivity, but do not formalize it by a marriage; and the man then has sexual intercourse with another woman; then the principle of mutual trust will lead us to identify that the essential features that make adultery wrong exist in this case as well, even though it does not fulfill the definition of adultery.

These identifications, made possible by a principled approach to ethics, are all made impossible by a rule-based approach. On a rule-based approach, we have a set of rules, one or more of which apply to adultery; e.g., the Ten Commandments, one of which is "thou shalt not commit adultery." When judging any specific case, the only relevant question is: has the rule been violated? I.e. has adultery occurred? All the rule requires to be applied is a clear definition of "adultery"; once we have that definition, no other question about the specific circumstances and context of a case can be relevant.

To the extent that a person thinks in terms of moral principles, the rule-based approach will seem intuitively wrong. In the case of adultery, it would seem clear that the definition of the concept "adultery" does not identify what is essential for moral judgment. This intuition is what lends Stroll's Wittgensteinian approach its seeming plausibility. While all cases of adultery do have some features in common, it is intuitively clear that there are differences between different cases of adultery—e.g. between those in which adul-

terers do or do not deceive their spouses—that, for the purpose of morally judging the action, are more essential than the common features.

In sum, while Stroll purports to argue against the use of principles in moral reasoning, it in fact his case works only as an argument against a morality based on rules. The argument's seeming plausibility comes precisely from the validity of a principled approach to ethics. Since ethical theories such as Utilitarianism and Kantianism do lead to rule-based moral reasoning, Stroll's argument does work against them; but the Aristotelian/Objectivist form of foundationalism, because it leads to thinking in principles, is not vulnerable to Stroll's argument.

### **Stroll's model of moral reasoning vs. principled reasoning**

Stroll proposes an alternative model of moral reasoning, which he claims to represent “the reasoning of ordinary, non-philosophical folk” in dealing with moral issues; and which philosophers need to adopt in order to “understand, accurately describe, and do justice to the facts of moral life” (Stroll, p. 320). Stroll's discussion and his examples do not discuss how people use morality to guide their own actions, and it is difficult to discern what his model implies about guiding one's own actions. Rather, Stroll focuses in his examples on how people make condemnatory judgments about other people's actions. While, as I note above, this is a very small part of the function of moral principles, it is still an area in which a principled approach is essentially different from a rule-based approach, as demonstrated in the example above of judging adultery. Below I examine Stroll's model of how such judgments are made, and contrast it with a principled approach.

In Stroll's model, people approach cases that need to be judged morally by comparing and contrasting them with other cases, examining their specific similarities and differences. Stroll provides a brief description of his model (pp. 315-318), but in it he never makes clear how such comparisons can lead to any moral judgments. However, he then follows with an illustration (pp. 318-320) of how his model would apply to a specific case, and the method he would suggest for reaching moral judgment is implicit in that discussion.

Stroll discusses a case in which a married couple adopted a daughter; after the adopted daughter had grown up, the husband had an affair with her, and later divorced his wife and married the adopted daughter. The question is: should the husband's behavior be condemned as incest? To answer the question, Stroll suggests that the case be compared to other cases which are or are not judged as incest. It is different from many other cases in that the adoptive father and daughter were not blood relations. However, there are cases in which sexual intercourse between two persons who are not blood relations is still condemned as incest; for example, in some tribes sexual intercourse between a brother-in-law and a sister-in-law is condemned as incestuous. There is also anthropological and historical evidence that sexual intercourse between blood relations is not a sufficient condition for incest; for example, in ancient Egypt, some members of the royal family were required to marry their own sisters, and this was not regarded as incest. All these are considerations which demonstrate that the lack of a blood relationship does not defend the adoptive father in our case against the accusation of incest.

The above illustration makes clear Stroll's model of how moral judgments are reached from examination of cases; they are reached by applying societal rules, which are accepted uncritically and unreflectively. The case is judged by observing how societies (whether contemporary American society or others, such as the society of ancient Egypt) apply the rule prohibiting incest, with no concern for why this rule should be respected at all or for the reasons societies had for regarding cases as covered or not covered by the rule. Stroll is thus accepting the essential feature of the ethical model he has argued against—morality as consisting of a set of rules; his disagreements are only in his rejection of any attempt to justify the rules, and in the method for deciding when the rule covers or does not cover a given case.

Stroll's claim, that his model reflects actual, real-life moral reasoning, can thus be stated more clearly as: "in real life, people uncritically and unreflectively accept societal moral rules." Is this claim correct? Clearly, it can't be universally correct, since then no society's dominant moral views would ever change. And even looking at those people who do accept their society's dominant morality, it is questionable whether all of them are as totally unreflective, lacking any concern for the justification of moral standards, as Stroll's model assumes. But most important, even to the extent that people are unreflective about societal moral standards, it does not follow at all that philosophers should accept the same unreflective attitude, by adopting Stroll's model; it is precisely the job of moral philosophers to provide the reflection that some non-philosophical people neglect.

Let us compare Stroll's example, of the application of his method, to a principled approach to the same case.

To take a principled approach, we have to identify a relevant principle, which can be justified based on man's basic needs and capacities. I submit that the principle relevant to this example is: "Sexual relationships should be voluntary on the part of both parties." This principle provides comprehensive guidance, identifying a wide range of wrong actions in many different situations: rape, child molestation, some cases of sexual harassment, as well as cases of incest. In many cases of incest, one of the parties—e.g. a daughter having sexual intercourse with her father—is subject to the authority and coercive power of the other party, to a degree seldom found in any free country outside of the family unit; because of that, such sexual relationships can't be regarded as voluntary, even if no direct use of force was involved. Also, in most cases of incest, one of the parties is too young to understand the nature and significance of the relationship, and is therefore incapable of giving informed consent to it; this, again, prevents the relationship from being voluntary. For these reasons, the principle justifies a condemnation of many cases of incest.

(This principle is not the historical reason for the rule against incest. That rule came to some extent from concerns about inbreeding leading to genetic deterioration—concerns that are irrelevant if the incestuous couple have no children—and to some extent from a desire to assert the authority of God by imposing arbitrary rules. With modern medical technology, when childbirth is no longer an unavoidable consequence of sexual intercourse, the principle I discuss here is the one remaining rational justification for the continuing condemnation of incest.)

By applying this principle, we can identify the essential questions that need to be answered to morally judge the specific case Stroll raises. The questions are: did the adoptive father, at the time the sexual affair began, possess authority and coercive power over his adoptive daughter, strong enough to prevent the relationship from being voluntary on her part? And was she mature enough to fully understand the nature and significance of the relationship? The principled approach thus makes it possible to identify those features of the case that are essential to judging it morally. Applying it to the specific case may still be difficult; it may be very hard to be certain of the answer to either of the above questions; but at least the principled approach allows us to identify the important questions to ask. That is what Stroll's approach does not make possible; no amount of comparing cases to each other would in itself help single out those similarities and differences that are essential to moral judgment.

Earlier, we saw that Stroll's Wittgensteinian argument gained its seeming plausibility precisely from the failure of rule-based ethics to identify the morally essential features of cases. We see here that Stroll's own suggested approach suffers from the same failure, and for the same reason—its reliance on rules. The argument Stroll has raised, against the Kantian or Utilitarian model he criticizes, therefore turns out to be an equally strong argument against his own approach to moral reasoning. In contrast, the Objectivist/Aristotelian form of moral foundationalism, by allowing a principled approach to ethics, completely disposes of Stroll's objections.

#### NOTES

- 1 Avrum Stroll, "Ethics Without Principles," in K. S. Johannessen, T. Nordenstam (eds.) *Wittgenstein and the Philosophy of Culture* (Vienna: Holder-Pichler-Tempsky, 1996.), pp. 310-320.

# Zoon Eleutherion: Man as an Individualist Animal

*Classical Individualism: The Supreme Importance of Each Human Being.* By Tibor Machan. *Studies in Social and Political Thought.* New York: Routledge, 1998. 243 pp. + xvii.

Irfan Khawaja

## I. Introduction

Tibor Machan's *Classical Individualism* (hereafter, *CI*) brings together a number of themes on the broad topic of "individualism" that have occupied the author's attention for the past three decades. The book consists of a Preface, with fifteen essays and an Epilogue, covering fairly abstract issues like free will and the objectivity of moral value, and relatively concrete ones, like multiculturalism and environmentalism. Like much of Machan's work, *CI* is informed in its broad approach by the writings of Ayn Rand, and Machan devotes a chapter at the end of the book to acknowledging and accounting for this intellectual debt (*CI*, ch. 15).

The breadth of Machan's arguments makes it impossible to discuss each of the essays in turn in the space at my disposal, or even to do justice to any significant number of them. Instead, I'll focus here on what I take to be two overarching themes of the book—the *metaphysical* basis of individualism in free will, and the *ethico-political* basis of individualism in a broadly Aristotelian conception of value.

## II. Machan on metaphysical freedom

It's a commonplace of philosophy that the term "freedom" is multiply ambiguous, having one group of senses in the context of metaphysical debates, and another in the context of debates about politics. On a fairly standard account of the distinction, *free will* pertains primarily to mental acts internal to the agent, which are beyond the power of others to control, while *political freedom* consists in the absence of restrictions on overt actions, which lie within the jurisdiction of government and can be violated by the acts of others. As a stylistic matter, I'll occasionally use the term *metaphysical freedom* as a synonym for "free will," and *metaphysical libertarianism* as a synonym for "the doctrine of free will."

It is, of course, important to distinguish metaphysical from political freedom as philosophers typically do, since the two are importantly different phenomena. The failure to distinguish between them leads to confusions like that of the hapless undergraduate I once taught, who, when asked to write an essay on "the topic of free will," instead produced an impassioned tract advocating

the repeal of all gun control laws, on the grounds that such laws constituted an impediment to “our freedom as Americans.” On the other hand, it’s possible to go to the reverse extreme, and fail to see that there is an important connection between political and metaphysical freedom—a connection justifying the thought that both things are related to the same phenomenon, the freedom of the individual from restrictions on moral action.

Machan forges a connection between metaphysical and political freedom by stressing the need for a *normative* argument for political freedom (*CI*, Preface, chs. 1, 2, 10, 13). The root meaning of the word “norm” is that of a principle or standard of evaluation for guiding action. Political freedom is a normative concept, Machan argues, because it denotes the absence of coercive restrictions over action to which we have a *right* (*CI*, p. 164). And rights, in turn, are principles whose primary function is:

to provide adult persons with a sphere of moral jurisdiction. This is due them because of their moral nature, because they have moral tasks in life that they ought to fulfill. Intruding on their sphere of moral jurisdiction would amount to thwarting their moral agency. And basic rights spell out where the conduct of others would or would not amount to intrusion. That is why the ‘border’ analogy is useful, even if it runs the risk of giving a physical image of a person’s sphere of moral authority. Moral agents require borders around them so as to know what their responsibilities are and where others must ultimately leave decisions up to them (*CI*, p. 122).

Rights, in other words, have the irreducibly normative function of identifying where it is that moral agents *must* be treated as *sovereign* over their actions. *Justice* tells us that within specified limits, they are *entitled* to assume control over their own actions without interference by others. The normativity of rights arises from the fact that the italicized terms in the preceding two sentences are ineliminably part of the concept of rights, and all four of those terms are normative. If *political freedom* is defined in terms of *rights*, and *coercion* is defined correlatively in terms of *political freedom*, then political freedom and coercion are normative terms as well.

The move from political to metaphysical freedom now becomes fairly straightforward. If political freedom is in fact morally valuable, and worth protecting, then respect for its conditions is a moral obligation. But if moral obligation entails moral responsibility, and responsibility entails metaphysical freedom, then an argument for political freedom must affirm free will and reject determinism (*CI*, ch. 2). Free will, in turn, requires an agent-causal conception of human action (*CI*, p. 23). So a commitment to political liberty ultimately turns out to be grounded in a metaphysical commitment to individualism via agent-causality.

Machan's central discussion of these claims comes in chapter 2 of the book, which lays out a case for free will, and discusses the problematic nature of its neglect in ethics, political philosophy, and the social sciences. The most valuable feature of the chapter is Machan's attack on John Rawls's "independence thesis," which denies the existence of a methodological priority between the claims of "comprehensive" doctrines on the one hand, and those of moral and political theory on the other.<sup>1</sup> A *political* theory of justice, on this view, must be "freestanding":

While we want a political conception to have a justification by reference to one or more comprehensive doctrines, it is neither presented as, nor is derived from, such a doctrine applied to the basic structure of society, as if this structure were simply another subject to which that doctrine applied . . . To use a current phrase, the political conception is a module, an essential constituent part, that fits into and can be supported by various reasonable comprehensive doctrines that endure in the society regulated by it. This means that it can be presented without saying, or knowing, or hazarding a conjecture about, what such doctrines it may belong to, or be supported by.<sup>2</sup>

As Rawls puts it, a political conception of liberalism and justice is "impartial" on such issues as the existence of God, the problem of universals, the truth-conditions of moral propositions, and the nature of free will and mind; it is consistent with any and all "reasonable" conceptions of these things.<sup>3</sup> It is "as far as possible, independent of the opposing and conflicting philosophical and religious doctrines that citizens affirm."<sup>4</sup> As Machan points out, the independence thesis (and the method of reflective equilibrium) predictably sanctions a theory that is shielded from criticism by those who reject the intuitive presuppositions and "settled convictions" of Rawls's epistemic constituency. Rawls's appeal to this constituency for the warrant of his theory relieves him of the need to argue for its *truth*, which in turn saves him from having to confront the charge that the theory is false. Correspondence-truth and falsehood are beside the point; the theory aims at a "practical political purpose" that is realized if it "makes our considered convictions more coherent," where the pronoun "our" refers to "those who accept the basic ideas of a constitutional regime," defined as Rawls defines it.

Rawls's treatment of free will and moral responsibility is a case in point. On the one hand, Rawls repeatedly tells us that his political conception of justice "disavows reliance" on any particular metaphysical view.<sup>5</sup> On the other hand, he finds himself addressing topics like moral responsibility that seem to require metaphysical treatment of some kind. His solution is to insist that what he says on such topics is somehow impartial between competing theories. But

this claim is hard to square with passages like the following from *A Theory of Justice*:

The assertion that a man deserves the superior character that enables him to make the effort to cultivate his abilities is equally problematic [as the assertion that he deserves his native endowments in the first place]; for his character depends in large part upon fortunate family and social circumstances for which he can claim no credit. The notion of desert seems not to apply to these cases.<sup>7</sup>

Rawls's view of character-formation clearly privileges determinism over free will. At the very least, it tells us that determinism has more explanatory power than metaphysical freedom in the explanation of character-formation, since the tiny (apparent) concession that Rawls makes to free will ("depends in *large part*") is compatible with his rejection of "the notion of desert" in character-formation. This decided preference for determinism shapes the structure of his theory at a fundamental level by means of the constraints that Rawls places on the concept of "desert," and by implication on "justice." It is hard to see how such assumptions can be "irrelevant" to "the structure and content of a political conception of justice," as Rawls implies they are.

Given Rawls's influence on contemporary moral theory, and the frequency with which he is cited as an authority in the literature, Machan argues plausibly that Rawls is indirectly responsible (so to speak) for the creation of a cottage industry in applied ethics that takes determinism as its point of departure, and churns out its consequences, sometimes with tragic-comic effect (see *CI* ch. 2 and references). The result is an artificial and self-perpetuating consensus in philosophy on the irrelevance of metaphysical freedom to practical life. Having marginalized free will from explanations of action from the start, theorists then insist that it is irrelevant, on the circular grounds that it doesn't figure in explanations of action. Having marginalized "unreasonable" comprehensive doctrines, we are told that the search for a true foundation for a comprehensive meta-ethics is "delusional."

Machan disagrees, and develops his own agent-causal conception of free will in chapter 3 of *CI*. His thesis here might be expressed in the slogan *self-determination explains moral individuation*, that is, the way in which a moral agent initiates and takes responsibility for his action explains the ways in which that agent is the individual he is, or has the moral identity he has. In brief, the argument is this. Non-human animals are certainly "individuals" in the sense of being individually separate, three-dimensional entities, and even in the sense of having their own distinct quasi-moral "characters" or "dispositions." Pet owners, zoo keepers, and naturalists insist—often vehemently—that some animals are "sweet-tempered" while others "have a bad temper," and some are "timid," while others are "bold." But the character of a non-human animal is entirely a function of nature and nurture; such animals don't *control*

*the making* of their characters by controlling the causal influence of nature and nurture on them. Non-human animals are not, in short, self-determining or self-responsible agents.

By contrast, human beings are. How a person controls the antecedent factors that influence his action is an essential part—if not *the* essential part—of the etiology of human action. Machan correctly stresses that self-determination is a *causal* variable—that it actually explains empirical events in the real world in a way that cannot, in principle, be explained in any other way. The importance of this point is hard to underestimate, since it is often ignored even by those who defend free will. Thus one advocate of free will tells us that though we have free will, it “must account for a fairly small percentage of the things we do.”<sup>10</sup> Another goes so far as to assert that:

If prediction and explanation are paradigmatic of scientific understanding, it appears that agent causation neither contributes to nor detracts from such understanding. Its contribution rather would be to our understanding of ourselves as moral agents.<sup>11</sup>

Both claims amount to the concession that free will has no empirical, explanatory, or practical significance. When it comes to explaining human action, we are best advised to call on the resources of determinism.

The trick to constructing a plausible theory of free will is to combine the thesis that we are self-determining agents with the thesis that our actions are part of the causal order, and involve goal-directed states of motivation. The attraction of determinism consists in the seeming ease with which it makes action intelligible by citing its causes: it often seems easier to understand how an act could be caused and motivated by deterministic antecedents than by self-determination without determinative antecedence. Given this, self-determination seems at first glance to lead to the absurdities of counter-causality or randomness. To avoid absurdity, the metaphysical libertarian needs to give an account of mental causation which shows in detail how and why mental action can be self-determined and produce physical action. Following Ayn Rand, Machan locates metaphysical freedom in the phenomenon of mental focus, and such related phenomena as attention, evasion, and the like. These phenomena are the right ones for solving the free will problem because they are *basic motivations*, involving both cognitive and conative elements; they make reasoned deliberation and action possible without themselves requiring prior deliberation or action.

Machan applies this framework to explain the actions of Rhoda Penmark, a character in the film *The Bad Seed*, depicted there as a quasi-psychopathic murderer. Provisionally setting aside the possibility that Rhoda is the victim of “involuntary cognitive impairment,”<sup>12</sup> Machan sets out to show, by the method of inference to the best explanation, that Rhoda’s actions are best explained by

appeal to intentional and volitional evil. The explanation involves three important variables. As a preliminary, Machan claims, Rhoda's actions can only be made intelligible if we see them as *intentional* or *volitional*, a fact that opens the door to an agent-causal explanation. Second, the relevantly explanatory volitions involve subversion of Rhoda's cognitive faculties best understood as a form of *voluntary self-subversion*. Finally, the self-subversion in question consists in an *evasion* of moral principles to which Rhoda (presumably)<sup>13</sup> had epistemic access, and thus serve to indict her of a form of irrationality that Machan calls "cognitive malpractice." Machan's explanation succeeds to the extent that these three variables in fact explain what Rhoda does better than rival explanations can.<sup>14</sup>

How well does Machan succeed? Not having seen the film, I find it hard to say; impressionistically, I found the account plausible but incomplete. But completeness is neither to be expected, nor the appropriate standard of evaluation for a single article on the subject. The value of Machan's account lies in its giving us an ingenious "just-so story" for explanations of intentional evil—one that invites further development. As James Lennox has argued in a different context, "just-so stories" function as tests of a theory's explanatory potential: they lay out the broad framework within which we can isolate the variables of a full explanation, plan a research program, and eventually test the results of that research.<sup>15</sup> By isolating volitional evasion as a factor in the explanation of evil, Machan opens up a number of exciting directions for future research at the frontiers of action theory and moral psychology. What makes that prospect especially exciting, I think, is the way in which Machan's account coheres with what we know about evil from other sources in the recent literature. Sometimes an explanation is incomplete because it is empirically inadequate. The coherence of Machan's account with that wider literature suggests just the opposite—that his theory's incompleteness is less a matter of empirical inadequacy than a sign of explanatory potential waiting to be realized.

### III. Machan on individualism

"Individualism," writes Susan Love Brown,

has been a dirty word ever since the French coined the term *individualisme* in the nineteenth century to label the horrific phenomenon that had overtaken their country in the form of a bloody revolution based on such radical ideas as individual rights and the rule of reason.<sup>17</sup>

In this respect, the provenance of "individualism" resembles that of "egoism" and "capitalism." All three terms were coined with the explicit intention of tarnishing each target by association with something obviously perverse, irrational, and evil. In each case, the coiners' strategy consisted in what Ayn Rand has aptly called "package-dealing": subsuming fundamentally dissimilar items under a single word, and then using that word univocally to convey the illusion

of similarity between the items.

The standard story in mainstream political theory holds that while “individualism” has some legitimate attractions, it is in most respects a “problematic” ideal. But so is collectivism. Therefore, we’re best off rejecting the “simplistic categories” of individualism versus collectivism in favor of a “more sophisticated” terminology that transcends them both.<sup>18</sup> This “sophisticated” insight is typically applied with a double standard, however, and theorists often forget the supposed attractions of individualism in their more moralistic moments, dredging up its Jacobin connotations when they need to make a polemical point. The result typically sounds something like this: “This book... describes a way of life that is dying—the culture of competitive individualism, which in its decadence has carried the logic of individualism to the extreme of a war of all against all, the pursuit of happiness to the dead end of a narcissistic preoccupation with the self.”<sup>19</sup>

As Machan points out, despite the lip service given to the ideals of individualism, it is precisely this diabolical conception of it that informs a good deal of contemporary political theory. Though theorists tell us that they are only interested in giving the concept some “nuance,” the nuance-giving is often led, as if by an invisible hand, to the task of delegitimizing individualism by packaging it with hedonism, narcissism, anomie, Social Darwinism, or Jacobinism. The plausibility of such maneuvers rests on a refusal to give the term a definition, but an insistence on using it to conjure up images of social or psychological dislocation.

Contemporary critics of individualism might be said to fall into five groups:

1. *Neo-Marxists* (e.g., C.B. MacPherson, Jon Elster);
2. *Conservatives* (e.g., Leo Strauss, John Gray, Irving Kristol);
3. *Communitarians* (e.g., Alasdair MacIntyre, Charles Taylor, Michael Sandel, Amitai Etzioni, Mary Midgely);
4. Various subgroupings of New Left *identity politics* (e.g., environmentalism, multiculturalism, some brands of feminism);
5. Non-Marxist *social democrats* (e.g., Richard Rorty, John Rawls, Martha Nussbaum, Will Kymlicka, Christopher Lasch).

Though the differences between these groups might matter in other contexts, such differences fade into the background when we view all five groups against the foil of Machan’s classical individualism. Whatever the differences between them, all five groups share an explicit hostility to the view Machan espouses. This includes groups (2) and (5), which contain members that in some sense might be called conservative or liberal “individualists.” None of these theorists, however, would be comfortable with Machan’s conception of individualism as individual sovereignty over the products of one’s labor, or of individual rights as restricting the government to the functions of a classical liberal state (*CI*, chs. 5, 10). A good deal of *CI* is devoted to responding to all five sorts of critics.

To this end, Machan distinguishes between two brands of individualism, Aristotelian and Hobbesian. *Hobbesian individualism*, on his account, is the problematic form, characterized by nominalism about universals, subjectivism about value, and atomism about human nature. *Aristotelian individualism* is the “classical” and defensible form, characterized by conceptualism about universals, objectivism about value, and what we might call biosocial essentialism (my term) about human nature. On this latter Aristotelian or classical version of individualism, Aristotelian individuals ought to be the primary unit of analysis in normative theory, and the primary concern of a legitimate social order. Each of us ought to strive, as Aristotelian individuals, to regard the pursuit of our own happiness as our overriding moral obligation. A just social order would respect that obligation by protecting the conditions that facilitated its optimal pursuit by each of us. Machan argues that the anti-individualists mentioned above are successful in their attacks on Hobbesian individualism, but fail to distinguish between it and Aristotelian individualism, which they leave entirely unscathed in their criticisms.<sup>20</sup>

Among the criticisms Machan works to overcome in *CI* is the objection that the very idea of “Aristotelian individualism” is incoherent. Aristotle, after all, is best known for his dictum that “man is by nature a political animal.” Anti-individualists have often used this Aristotelian thesis to argue against individualism as follows: 1) Aristotle was correct to argue that humans are by nature political animals; 2) but individualism denies this Aristotelian truth; 3) hence individualism is false. The argument raises a dilemma for Machan: if classical individualism is Aristotelian, it can’t be genuinely individualistic; but if it’s really individualistic, it can’t be genuinely Aristotelian. So, the criticism goes, Machan must choose between his commitments to Aristotelianism and to individualism.

Machan, however, believes that he can have both Aristotelianism and individualism simultaneously. Granting the existence of contrary evidence, he isolates a solid core of textual evidence for a form of individualism in Aristotle and generally in the Aristotelian tradition. The plausibility of Machan’s argument derives from the fact that individualism is in fact a pervasive theme in several important elements of Aristotle’s philosophy. Thus some support for individualism comes from Aristotle’s metaphysics of entities which, to quote Edward Zeller, makes “the Individual...the primary reality” in Aristotle’s ontology, and gives it “first claim on recognition” (*CI*, p. 175). Some of it comes from Aristotle’s theory of action, which is the *locus classicus* of the agent-causal theory of free will that Machan defends elsewhere in the book. Some of it comes from Aristotle’s theory of value, which makes an individual organism’s flourishing that organism’s ultimate end, and the source of the norms that guide its life. Some of it comes from Aristotle’s theory of practical reasoning and virtue, which places a high premium on ordering one’s life by one’s own rational choices. Some of it even comes from the most anti-individualist part of Aristotle’s philosophy, his politics; in a justly-celebrated study, Fred D. Miller Jr. has recently argued that Aristotle’s political theory gives a central place to individual rights and a “moderately individualistic” theory of the common good.<sup>21</sup> Machan usefully points to similarities between this Aristotelian conception of

individualism and various historical influences on contemporary life, from Christian and Islamic theology, to classical liberalism, to the thought of the American Founders, to the writings of Ayn Rand (*CI*, Preface, chs. 1, 14, 15).

One of the virtues of Machan's discussion is that he manages to maintain a healthy sense of perspective on the texts, making a good case for Aristotelian individualism while acknowledging the existence of other ways of reading the texts, and some texts that contradict his interpretation. The purpose of appealing to the texts is to identify two forms of individualism at a fairly high level of generality, and the evidence that Machan cites is more or less sufficient for this task. In this respect, Machan's approach differs drastically from some of his critics, whose *modus operandi* consists in making bold, unsupported—and occasionally downright wild—assertions about the relationship between Aristotle and individualism. A close reading of the Preface, and of chapters 1, 4, 14, and 15 of *CI* should give such critics pause, and give others a lot to think about.

Having made the case for the coherence of an Aristotelian form of individualism, however, it's a separate task to make that case relevant to contemporary life. Aristotle lived nearly 2400 years ago in a slave-owning, deeply misogynistic society, and explicitly deprecated the value of productive work. In fact, Aristotle's view of productive work—that it is an inferior task performed by inferior people whose products can be expropriated at will (cf. *Politics* 1254a4-8)—is not only the antithesis of Machan's individualism, but is arguably one of the sources of opposition to it. Drawing on Locke and the other classical liberals, Machan works to detach these Aristotelian prejudices from Aristotle's more fundamental claims (e.g., those mentioned above), and then connects these fundamental claims with an essentially Lockean politics. One of the best results of this approach is Machan's treatment of the so-called "tragedy of the commons," which he renames the *moral tragedy of the commons*, and conceptualizes in a way that is both clearer and deeper than that of its "original" author, Garrett Hardin (*CI*, p. 49). The idea of a moral tragedy of the commons has deep roots in Aristotle's critique of Platonic communism, and in Locke's theory of property; Machan redeploys the concept to offer criticisms of redistribution and environmentalism that maintain continuity with the Aristotelian and Lockean arguments (*CI*, chs. 5, 10, 11, 12, *passim*).

My one objection to Machan's treatment of individualism is his tendency to overstate the case against "atomism" without really making clear what is supposed to be wrong with it. In fact, Machan goes so far as to assert that human beings are "essentially social animals," and (I assume) takes "atomism" to be the denial of this claim.

This strikes me as too much of a concession. It is, I think, far from obvious that human beings are "essentially social animals" in the literal and technical sense that "sociality" is the essence of the animal *Homo sapiens sapiens*. For one thing, that assertion would require a worked-out theory of essentiality, which neither Machan nor his anti-individualist critics discuss. It would also require criteria for what it is to be a social *animal*—criteria that do not yield

obvious results in the human case. On one definition, a “social animal” might be an animal species whose adult members were literally incapable of physiological self-maintenance apart from a social structure of other conspecific adults. Honeybees and ants seem to fit this description, but even intuitively “social animals” like lions do not. Though lions usually live in groups—and usually enhance their life-prospects by doing so—adult lions promote their *own* survival as individuals (not that of their prides, much less their species), and are capable of surviving as nomads. In fact, in some cases nomadic life is *better* for a given lion than life in a pride.<sup>22</sup>

Human beings may in some sense be “social animals,” but if so, they are more like lions than they are like honeybees or ants. In fact, focusing on our unique attributes, it is clear that our life-functions are far *more* individualistic than those of lions or of any non-rational animal. If flourishing is a matter of self-generated and self-sustaining action, then an animal’s essential attribute is the irreducible capacity that is causally responsible for generating, directing, and sustaining action in that kind of animal. In the human case, that attribute is *reason*, which is causally responsible for our acquiring knowledge, setting ends, and determining action. To be sure, social life plays an important role in the development of that faculty, and in creating the conditions for its optimal functioning. But that only means that sociality is important to human life, not that sociality is our essence. If we have free will, we initiate action on our own; if knowledge is a property of individuals, we acquire it on our own; if survival qua human is our ultimate end, we pursue it for ourselves; if we are sovereign agents, then our persons constitute inviolable boundaries against others; if self-esteem is a value, we can enjoy it by and for ourselves.

We identify the requirements of the best society in terms of *these* needs, individualistically conceived, not vice versa. That is why, as Ayn Rand puts it, “No society can be of value to man’s life if the price is the surrender of his right to his life.” Notice that Rand’s claim presupposes that we can identify the requirements of “man’s life” prior to and independently of our conception of human sociality. The requirements thus conceived provide the standard for a *subsequent* conception of human sociality. We can’t, without vicious circularity, begin with an irreducible conception of “sociality” and then assume that “man’s life” includes it as a constituent.

If so, I don’t think that there’s much to be gained by denying that classical individualism is “atomistic.” It’s worth remembering that atoms form bonds based on their natures—powerful and lasting ones. It’s hardly an indictment of a naturalistic theory of individualism to recognize that, in our own way, we do the same thing—with the difference, to paraphrase Thomas Jefferson, that when the need arises, we can choose to dissolve our social bonds in the name of independence.

#### **IV. Conclusion**

Though virtually every reader of *CI* will have quarrels and quibbles with

Machan's claims, as I did, most readers, I think, will find it a useful distillation of the case for neo-Aristotelian individualism, and a catalyst for more finely-grained work on the issues it raises. Action theorists, ethicists, and theorists of culture might pay attention to the details of Machan's account of volitional evil, and test the plausibility of his just-so story on works of fiction, history, journalism, or on everyday life. Social scientists and political philosophers might ask how the concept of "a moral tragedy of the commons" integrates causal consequences and normative principles in the case for classical liberalism. And historians might ask questions about the explanatory power of the distinction between Aristotelian and Hobbesian individualism, whether applied to texts or to events. There is much more in the book than can be conveyed here, however; anyone interested in individualism would profit by taking a look at *CI* for himself.

## NOTES

1. The initial statement of the thesis is in John Rawls, "The Independence of Moral Theory," *Proceedings and Addresses of the American Philosophical Association*, vol. 49 (1974). Machan mostly limits himself to Rawls's earlier statements.
2. John Rawls, *Political Liberalism*, paperback edition (New York: Columbia University Press, 1996), p. 12.
3. Rawls, *Political Liberalism*, pp. 58-66. The point is hard to state with any precision, since Rawls tells us that his "account of reasonable comprehensive doctrines is deliberately loose" (p. 59). The 1974 statement of the view (note 1) is clearer.
4. Rawls, *Political Liberalism*, p. 9.
5. The quoted phrases are all from Rawls, *Political Liberalism*: "settled convictions" (p. 8), the remainder from p. 156.
6. The quoted phrase and one instance of the claim are found at *Political Liberalism*, p. 29. One of the puzzles of this "disavowal" is that Rawls concedes that he is unsure of what is being disavowed (p. 29n.31). He then ventures some suggestions, "avoids" commenting on them, and finally offers an alternative tentative suggestion without fully endorsing it.
7. John Rawls, *A Theory of Justice*, (Cambridge, MA: Harvard University Press, 1971), p. 104, discussed on p. 22 of *CI*. Rawls asserts the basic continuity of *A Theory of Justice with Political Liberalism* on p. xvi of the latter book.
8. Rawls, *Political Liberalism*, p. 29n.31.
9. Rawls, *Political Liberalism*, p. xviii.
10. Peter van Inwagen, "When is the Will Free?" in *Agents, Causes, and Events: Essays on Indeterminism and Free Will*, edited by Timothy O'Connor, (Oxford: Oxford University Press, 1995), p. 235.
11. Randolph Clarke, "Toward a Credible Agent-Causal Account of Free Will," in *Agents, Causes, and Events*, p. 210.
12. Machan sets it aside not necessarily because he takes it to be false, but because he wants to offer a disjunctive explanation: *either* Rhoda is involuntarily impaired—or if she is *not* involuntarily impaired, then she is volitionally evil. Obviously, since we're dealing with a fictional character in a film, Machan can't have access to all of the information needed to offer the "true" explanation of Rhoda's actions, so the disjunctive procedure is perfectly justified on his part.
13. Again, subject to the proviso in the preceding footnote. If Rhoda were involuntarily impaired, she could not have had epistemic access, but if she were not involuntarily impaired, she would have had to evade her own moral knowledge.
14. For a rival explanation, see Laurie Calhoun, "Moral Blindness and Moral Responsibility: What Can We Learn from Rhoda Penmark?" *Journal of Applied Philosophy*, vol. 13 (1996), pp. 41-50. Calhoun's essay contains more detailed analysis of the film than Machan's, but suffers from what seems to me a serious difficulty. On the one hand, Calhoun argues that Rhoda "has no cognitive deficiencies whatsoever" (p. 41); on the other hand, Rhoda is described as "morally blind." On the face of it, this seems self-contradictory: "moral blindness" is typically taken to be a failure of moral discernment, which is precisely a "cognitive deficiency." Further, Calhoun describes Rhoda as engaging in "rationalization" (p. 43), and as being self-deceived (p. 47). Calhoun's response seems to be that the relevant explanatory variable is lack of the relevant moral (but not cognitive) sentiments (p. 49). But this explanation flies in the face of Calhoun's own evidence, which bears on cognition.
15. James Lennox, "Darwinian Thought-Experiments: A Function for Just-So Stories," in *Thought Experiments in*

*Science and Philosophy*, edited by Tamara Horowitz and Gerald J. Massey, (Savage, MD: Rowman and Littlefield, 1991), pp. 236-238.

16. The most famous source is Hannah Arendt; see *Eichmann in Jerusalem: A Report on the Banality of Evil* (New York: Viking, 1964), pp. 287-88. See also Neera Badhwar, "Self-Interest and Virtue," in *Self-Interest*, edited by Jeffrey Paul, Fred D. Miller Jr., and Ellen Frankel Paul, (Cambridge: Cambridge University Press, 1992), pp. 226-263, and David Kelley, "Stalking the Criminal Mind," *Harper's* (August 1985), pp. 53-59.
17. Susan Love Brown, "Breaking the Habits of the Heart," *Critical Review* vol. 5:3 (Summer 1991), p. 379.
18. By "mainstream," I mean: "liberal academic theorizing in Departments of Politics and Philosophy in the English-speaking world, usually influenced primarily by Rawls and/or Dworkin." Non-mainstream leftists whose views are to the left of the mainstream consensus are often entirely comfortable with the individualist-collectivist dichotomy. See, for example, Robin Blackburn, "The New Collectivism: Pension Reform, Grey Capitalism, and Complex Socialism," *New Left Review*, vol. 233 (January/February 1999), pp. 3-65. Blackburn's purpose in the article is explicitly to defend a form of collectivism against individualism. One seldom sees mainstream liberals inveighing against the "simplistic" nature of this use of the dichotomy, however.
19. Christopher Lasch, *The Culture of Narcissism*, (New York: W.W. Norton, 1978), p. xv.
20. For a concise statement of Aristotelian individualism, see *CI*, p. 170.
21. On action theory, see Roderick Long, "Aristotle's Conception of Freedom," *Review of Metaphysics* vol. 49:4 (June 1996), pp. 801-2. On value theory, see Allan Gotthelf, "Aristotle's Conception of Final Causality," p. 233 in *Philosophical Issues in Aristotle's Biology*, (Cambridge, UK: Cambridge University Press, 1987), edited by Allan Gotthelf and James G. Lennox. On choice, see John M. Cooper, *Reason and Human Good in Aristotle*, (Indianapolis: Hackett Publishing, 1975), ch. 2. On politics, see Fred D. Miller Jr., *Nature, Justice and Rights in Aristotle's Politics*, (Oxford: Oxford University Press, 1995).
22. Anne E. Pusey and Craig Packer, "Divided We Fall: Cooperation Among Lions," *Scientific American*, vol. 276 (May 1997), pp. 52-59. I thank James Lennox for helpful conversation on the issues in this paragraph (but absolve him of responsibility for what I say here).

# Attending to Mind Itself

*Mind Regained* by Edward Pols.  
Ithaca, NY: Cornell University Press, 1998

Roger E. Bissell

With the publication of Edward Pols's recent books on the nature of free will (*The Acts of Our Being*, 1982, Univ. of Mass. Press) and knowledge (*Radical Realism*, 1992, Cornell Univ. Press), the components were in place for a comprehensive essay of the mind-body problem. It is no surprise, then, that Pols's newest book, *Mind Regained* (1998, Cornell Univ. Press), undertakes just such a study, continuing and intensifying his defense of the mind as being able to know reality directly and as being as real as the things it is able to directly know. Like Pols's previous works, *Mind Regained* is so rich in interrelated content that it defies brief, simple summarizing, so this review will do the next best thing and provide a lengthy, complex summary. As the barest of synopses, however: in the first chapter, Pols makes it clear that, whatever its limitations, the tradition based on Plato and Aristotle was basically right on the issues of knowing, causality, mind, and soul. In three succeeding chapters, he shows how modern philosophy, in its union with science, oversimplified and distorted these issues. Finally, in the last two chapters, he gives us the perspectives and techniques for setting everything right again.

There is a point to the hyperbole of the title chosen for this book: although the power of mind's functions has accomplished great things in science and technology during this century, Pols says, "something profoundly important to mind's well-being has indeed been lost, and lost by the very persons who should have been most zealous to preserve it—I mean the most influential workers in academic philosophy, cognitive science, and neurophysiology." These people have failed to see that mind "operates as a real cause within and upon the material world, and that this causality is the source of all the theoretical and physical devices" that allowed the revolutionary advances we have witnessed in our time; they have "lost an adequate understanding of the very functions by virtue of which [mind] accomplishes both its everyday and its more exalted tasks" (vii). Since the intellectual leaders have so egregiously dropped the ball, Pols aims his book not at them, but instead at the wider intellectual community. While finding the mind-body problem interesting and important, mainstream intellectuals have been so negatively influenced by the writings of talented establishment thinkers that they are "unwilling or unable to look at their own minds in action and find there what has been left out of that establishment account." Pols hopes to convince this wider intellectual public to engage in what he calls "attending to mind itself," in order to gain "a new self-

consciousness on the part of mind as it manifests itself in their persons.” He concedes that while attending to mind itself is not easy, it is “essential to regaining what has been lost, and . . . it is by no means a matter that must be left to professional philosophers and professional students of the mind” (viii-ix).

But why, specifically, has there been such a widespread loss of an adequate understanding of the mental functions by which we accomplish both our everyday and our more exalted tasks? In seeking an explanation, Pols focuses in the first part (fully two-thirds) of his book, “Attending to Doctrines,” on what he sees as the two most important factors: the predisposition of most contemporary philosophers to what he calls the “negative philosophical judgment about the powers of the human mind” and the prevalence among philosophers and scientists of the “received scientific doctrine of causality” (viii). The former factor, based on “the myth of the self-enclosure of the mind” (19), a perspective on human knowledge established by Descartes’s representationalism, and transmitted from Kant on down to the present, presents mind as having “no direct reality-attaining function [and having to] make do instead with a groping and always-frustrated approach to the real that begins in ideas and then makes its way by constructing theories about that which is not directly accessible.” The latter factor is the attempt, characteristic of 20th century reductionist philosophers and scientists, to portray mind as being less real than the physical infrastructure that supports it, to explain mind as “an *effect* of the physical entities that science investigates so superbly,” and to reject any theory that does not explain “any *apparent* causality on the part of mind . . . in terms of the doctrine of causality that prevails in science” (viii).

Notwithstanding the brilliance and usefulness of many recent achievements in the study of mind, Pols says, they were developed within a framework unnecessarily constricted by, and serving to perpetuate, the failure to grasp what was long ago acknowledged by the great philosophical doctrines of the ancient Greeks and medieval Scholastics: that “mind itself is the deepest ordering principle of nature or at least the most important expression of that ordering principle” and that “causality is hierarchical and that mind is central to that hierarchy” (1, 19). In guiding us back to such an understanding, albeit updated in content and method, Pols in the second part of his book, “Attending to Mind Itself,” employs ideas and technique that are both simple and profound. He invites us, in the spirit of Plato and Aristotle, to use the “reality-attaining powers” of mind and “approach the study of mind by way of mind itself rather than by way of its infrastructure,” and he shows us how to do so in a manner that avoids the disastrous errors engendered by Descartes in his own heroic efforts. In choosing to focus on the powers and functions of the mind itself, however, Pols in no way means to deny or dismiss the importance and massiveness of the contribution of the infrastructure in all the things that mind does, he merely intends to establish that “the physical basis of mind is not the only causal factor in mind,” that something causally significant, “some truly *causal* factor is missing when the study of mind is approached only by way of the central nervous system.” He further readily concedes that he is entitled to refer to the brain as belonging to “the infrastructure of mind” only if he is also

entitled to also refer to “the mind itself,” i.e., only if he is also able to make his case that there is more of causal significance in the mind’s functioning than is found by studying the nervous system (10-1).

If he is correct, Pols says, the implication is clear: human beings and their acts are irreducible “causal hierarchies.” He is thus also inviting us to consider and apply a model of causality which, though explicitly designed to be able to incorporate the advances made by modern science in understanding the infrastructure that supports the functions of mind, is more similar to that of Aristotle and Plato than that of Hume and Laplace. He seeks to convince us that rational action, which he calls a “master function within which we can discern other functions brought together under the telos [end, purpose] that defines the action itself,” cannot be completely analyzed and understood, from either a functional/temporal or a structural/spatial perspective, “as entirely an effect of causes other than itself.”

(1) Functionally, any attempt at linear-event analysis of an action into a series of mental or physical events is futile, in that it cannot account fully for the holistic unity, the wholeness, of the action. The reason, Pols says, is that an action is not “temporally linear,” instead requiring a non-discrete, “global” amount of time to be the action it is: “[T]he earlier ‘parts’ anticipate the later ‘parts’; the later ‘parts’ retain the earlier ‘parts’ in order to complete what was begun there.” The very nature of a rational act, with its “telic drive,” thus requires that its causality be not linear and one-directional, but (in a sense) circular and bi-directional. On the level of conscious purpose, anticipation and remembering are the functions that allow a writer, speaker, or performer to construct sentences or musical phrases in which an earlier word or note does not *cause* a later word or note, yet is selected in anticipation of its being an appropriate predecessor, with the latter being selected as an appropriate successor by the guidance of remembrance of what it is succeeding. Something like anticipation and remembering, though often not on the conscious level, are essential features of any purposeful act, the purpose of the overall act itself being a vital and central facet, if not the totality, of the cause that guides the quasi-anticipatory and recollective functions by which the various events characterizing the act are carried out (15-6, 19).

(2) Structurally, any attempt at analysis of an action into a series of infrastructure events again leaves the action as a whole incompletely accounted for: “The causal contributions of discrete infrastructure elements are no doubt real enough, but they do not appear as such in the act. They seem rather to be *used* by the act [and] absorbed into its temporal unity.” Though these elements are “accessible for study, . . . they are not accessible *as supporting mind*. It is only by deploying mind itself in the theoretical activity we call science that we can learn about that support . . . [W]e know well how to use neurons and so also electrons, even though we can give no account of that “how”: we need only deploy mind itself in whatever task or problem happens to interest us. If the infrastructure is healthy, it will support that activity of mind itself . . . [I]n the doing of the activity only mind itself is manifest to us” (14). Rather than being

caused by what goes on within its infrastructure, then, a rational act is *self-caused*, in the sense that “it makes use of the units of the infrastructure by incorporating them into its own actuality.” This is the key Pols offers for understanding how knowledge, values, and motives have causal significance in human action. Much of the “actuality” of a rational act consists of:

what mind comes to know in the course of the development of the act—the reasons it understands and assents to, the things it understands to be good and therefore to be pursued, the things it takes to be bad and therefore to be avoided. Things thus known are causes of the action in the limited sense that knowing is part of the action and these are things known. (16)

Pols begins in chapter 1, “Plato and Aristotle on Mind, Soul, and Causality,” with a survey of the relevant thought of these two “most influential representatives [of] the ancient hierarchical view of causality.” He first discusses the central feature of Plato’s philosophy, his doctrine of the Forms, which are “*what* mind knows *when* it knows, but . . . transcend the mind as well as all particular things . . . not abstracted from particular things: they simply *are* in their own right, and . . . provide . . . the being in which physical things merely participate” (23). Plato held that the soul’s purpose is to “bring life to the body . . . to rule or master the body and by doing so produce a complex of soul and body that is virtuous—in short, a good person” and, as a “theoretic mind,” to know the Forms (31). In explaining how the soul was to master the body, Plato proposed the idea that the soul was in three parts, the appetitive, the spirited, and the rational. The soul is a self-mover, being “moved by desire, or love, of the Forms and in particular the Form of the Good; and as a self-mover it brings life to the body (or the many bodies) with which it is joined for a time” (39). Plato further viewed “mind, or reason, [as] the highest functional level of the soul,” and as something that needs no causal explanation, for it instead is a powerful cause “that must be invoked in any adequate explanation of other things” (32-4).

Aristotle radically shifted the focus of the study of being. He regarded primary being not as eternal, unchanging forms considered in themselves, but rather human beings and other “particular, individual, changing, and developing being[s],” which he called “*ousia*.” What makes an “individual being a *being*, i.e., real or actual, Aristotle said, is form, which Pols redubs “immanent form/essence” and, equivalently, “informing form or essence.” Form operates concretely as informing form to make something an individual concrete being; form is “less than fully real—less than *actual* [i.e., immanent] . . . when it is not operating in that concrete way.” Informing form can be defined (e.g., man is ‘rational animal’), is not particular (i.e., is not the form ‘Socrates’), and it is an ideal (individuals can fall short of it, in the course of their lives); it also has “causal significance, for it is the reason why the individual being develops into a man by a long process of change—change that does not affect the form itself, which . . . merely is . . .” Aristotle distinguished form in this primary sense

from that same form as “abstracted by a knower and thought about, in which case it is no longer the causally vital form . . . of some particular being but rather a universal form . . .” As Pols points out, “when Aristotle wishes to single out the abstract form, he often calls it a universal.” Bearing this in mind will help us avoid the confusion that often results because “essence” is “sometimes applied both to the informing form in its concrete actuality and to the same form as entertained and defined by our intelligence (the abstracted form)” (35-8)

As known, i.e., in abstraction from the individual being that it informs, a form is not fully actual. For this reason, Aristotle refers to an individual as a “primary *ousia*” and the abstract (universal) form as a “secondary *ousia*.” (The medievals referred to them as “real beings” and “beings of reason,” respectively.) Moreover, since it is the informing form that makes something “a real primary being . . . Aristotle calls this informing form the actual *ousia*,” which is, however, “itself fully real (actual) only when it is [making] concrete things real.” Form as abstract/universal, on the other hand, “does not do or accomplish anything, and so it is secondary.” The secondary *ousia* is what we consider when we want to “define or otherwise consider rationally this vital source of the being of a primary *ousia*, and if we wish to take account of the fact that although it is the vital source of individuality, it is not itself an individual [and we therefore] must perform an act of abstraction.” By abstracting the form of something, we are considering that thing “in general,” i.e., as a universal, as opposed to considering that same thing in its concrete actuality, i.e., as a particular, made fully real by being informed by its form (37-8). Pols summarizes:

[T]he very actuality of a being is *form*; the active informing principle which makes it a concrete and functioning being worthy of the expression *ousia energeia* is also that which makes the being intelligible and thus capable of being defined . . . . [T]he form that informs a primary *ousia* and indeed makes it actual and thus primary also endows it with intelligibility and so makes it capable of providing to the inquiring reason that secondary *ousia* we call a universal or abstract form. (38)

Aristotle applied these insights first and foremost in his view of the soul, which he regarded as the actual, informing form/essence of a living organism, giving it both its *being*, i.e., its life, its movement and activity, on the one hand, and its intelligibility and definability on the other. In contrast to Plato, Aristotle considered the soul to be an “unmoved mover . . . of the body, providing the source of, and the goal for, the body’s movement and development.” In this respect, Aristotle considers the soul to be the formal, final, and efficient causes of the body (though not its material cause). Pols notes that Aristotle’s own tripartite model of the soul differs from Plato’s in that its first two parts (rational and sensitive/appetitive) correspond approximately to the three parts of the Platonic soul, adding a third part (nutritive/vegetative) not present in Plato’s

model. Choice, which Aristotle calls “reason that desires or desire that reasons,” is similar to Plato’s “spirited” part of the soul, in that it can perfect itself by allying with reason or corrupt itself by allying with appetite/desire. It is moderation of one’s appetites/desires that provides the conditions for the ideal, the rational, contemplative life. As in Plato’s model, each of the parts of the soul is, for Aristotle, a vital functional level of the whole rational animal, which has to function together with the others in an appropriate manner in order for one to live well, the lower levels serving as infrastructure for the mind on the top level of the hierarchy (39-42).

In chapter 2, “Descartes’ Dualism and Its Disastrous Consequences,” Pols highlights the main points and rationale of Descartes’s extreme mind-body dualism, with special attention to the aspects of it that led to the negative view of mind’s power to know reality and to the state-event model of causality. Descartes saw mind as so radically different from body that by the very extreme difference in their constituents (soul or “thinking substance” and matter or “extended substance”) neither could have a causal effect on the other. Despite his attempts to argue that the mind/soul and a machine-like physical body interacted via the pineal gland (!), neither the rationalist nor the empiricist philosophers who followed Descartes would accept the idea of mind-body interaction, instead opting for some form of parallelism or preestablished harmony, on the one hand, or materialistic monism, on the other. The empiricists leaned toward materialism because they uncritically adopted Descartes’s idea that “the reality of the physical world is different from what common sense takes it to be,” that the physical world is real while our common sense experience of it is not—or that the physical world is more real than our experience of it. Like Descartes, the empiricists believed that “knowledge does not consist in a relation between our mind and things but rather in the relation between the mind and its ideas . . .” Unlike Descartes, whose rationalist view that innate representative ideas gave the mind the ability to access the reality behind the veil of appearance, the empiricists held that the real source of knowledge was specifically the *impressions* in the stream of experience, which are not representative in a way that allows us to infer a reality beyond them, although prudence leads us to anticipate in their future course as much as possible. This effectively blocked empiricists from using experience as the basis for demonstrating that scientists are describing a real material world behind the appearances of the impressions of experience (54-5). They and the rest of modern philosophy remain trapped in what Pols calls “the central predicament of all post-Cartesian epistemology,” an unreal one to be sure, else (despite Descartes’s intricate and ingenious efforts) there would be no way out of it:

[I]deas purport to represent real things or features of real things, but we have access only to the representations and not to their originals, if indeed there are originals . . . Thus if subjects called bodies really exist—both the bodies that seem to make up the commonsense world and our own particular bodies, considered as parts

of that world—we can know what bodies are (their true natures) and *whether* bodies really exist independently of mind only by undertaking a demonstration that begins with the representative reality of our own ideas. (50)

Pols points out a further way in which Descartes's dualism has, more indirectly, undermined or distorted the empirical study of mind, which often adopts the assumption that "mind and body may be understood in terms of two distinct streams of states/events, one mental and the other physical." The idea of a succession of mental states/events is implicit in Descartes' idea that mind is a radically different substance from body, but it is developed in explicit detail in Spinoza's parallelism and, more important, in Hume's phenomenalism. Hume held that we cannot demonstrate the existence of the external world from the sequence of impressions in our experience. Yet, since he held that our knowledge of the mental stream is more certain than that of the physical stream, a dilemma arises. Hume wants to argue that the latter stream of events is more basic and that it is ruled by cause-and-effect; however, he also wants to argue that the existence of the physical world is a postulation based on the mental stream, and that causality is not based on necessities in the (inaccessible) external world but merely on the observed regularities of sequences of events in the mental stream. Strictly speaking, then, Hume's view limits empiricists to viewing causality in either stream of events as being mere "constant correlation: x is the cause of y if and only if when x occurs y follows and when x does not occur y does not occur." Nor is there any apparent solution to the problem of which stream is more real or basic. The hard-nosed realism of the causal views of mainstream empiricist materialism are at sixes and nines with the watered-down causality of Hume's phenomenalism (56-7).

Pols follows up on this problem at the end of chapter 4 by briefly pointing out that the difficulties with conceiving of causality exclusively in terms of the succession of physical states and events have a parallel in the error of arguing that the powers of the mind are exercised only by means of a stream of mental states and events. The attempt of some contemporary analytic-empiricist philosophers to thus "assimilate a supposed mental causality to the received scientific doctrine of causality" is not an adequate correction to the materialist emphasis on the physical. Such mental states and events are, in fact, "abstractions from the lives of persons. In all plausible cases of what at least purports to be causally significant mental activity, it is only after someone has acted rationally that you can pick out with any confidence a series of states/events (of whatever kind) and consider their causal role in certain purposive achievements" (91). What this abstract state/event model, in both its physical and mental variants, leaves out of an explanation of rational action and purposive achievement is the "telic unity" of their temporal structure, i.e., "the directed unity of [their] several stages." (92). A comment in the following chapter puts an appropriate cap on this point:

Given the profusion of acts of the mind that are intricately and ineluctably embodied, the notion that mind can be adequately described in terms of sequences of purely mental events set in contrast with sequences of purely physical events taking place in the brain seems an unreal contrivance. The contrivance is based on the obsessive notion that any physical event singled out from a physical system is wholly caused by prior physical events. Take that notion and apply it to supposed mental events and you have a straw-man dualism that can then be easily discredited in favor of one of the many forms of physicalist monism that are current today. (100)

Most of Pols's concern in chapters 3 and 4 is to reveal the problems with conceiving of causality exclusively in terms of the succession of physical states and events. One of the chief problems is that the very prestige of this viewpoint inhibits many who study the mind from trying to see whether the physical world in general and human beings in particular exhibit hierarchical causality. In chapter 3, "The Received Scientific Doctrine of Causality," Pols traces the historical process by which Aristotle's "four causes" were gradually replaced. The scientific efforts of Kepler and Galileo led to the modern view of causality that strips away formal causality in the full sense and thus telic or final causality as well. The received scientific doctrine also reduces material causality from the idea that "the hierarchical principle by virtue of which what was a formal cause at one ontological level could serve as a material cause for a higher (formal) ontological level" to the idea that inferred microentities are "more truly real than the entities to whose macroscopic structure they contribute, . . . the observed forms of macroscopic entities [being] dependent on the observer in a way analogous to such secondary qualities as colors;" reduces formal causality from the idea that a visible, intelligible structure emerges from a process of change to the idea that a law of nature (e.g., the laws of Newtonian mechanics) displays the mathematical form by which atoms and larger entities move; and treats efficient causality as interaction between observed entities, which move the way they do because they are composed of atoms (64-5). Following Newton's acceptance of atomism, Laplace assumes that nature is really a concrete physical system composed of "microscopic particles moving inexorably from one state to another and giving rise to all the macroscopic realities to which human beings respond." Laplace's view of causality thus sees the universe as a whole as being "a physical system that passes through successive *states*, any given state being the *cause* of the state that follows;" and "the transition from state to state is governed by laws of nature . . . [T]he laws of nature are causal factors no less than the physical states are" (69).

Since most research and applied science focuses on physical systems smaller than the universe as a whole, the "working model," the physical sys-

tems model, for Laplace's view of causality regards a physical system at some particular time as the cause and that same system at a later time as the effect. Further, while this kind of mathematical analysis of two sequential states of the same physical system is sometimes most appropriate, at other times all that is necessary is a simpler model of causality that links two particular entities/events as the cause and effect of some transaction, the first event, condition, or entity being the cause of the second only if it is "necessary and sufficient" for the second (74). In an effort to precisely define "sufficiency," philosophers tend to argue in terms of a given transaction being governed by law. There is, however, a deep split among scientists and philosophers about the ontological status of the laws of nature. The Laplacean optimists because of their confidence in the realism of laws and their own ability to know things as they are, view the laws of nature as "prescriptive in a causally determinative sense . . . rather than merely descriptive . . . . [W]hat is thus explained could in principle have been predicted" (76). The Humean pessimists on the other hand argue that "there is no justification for the claim that we as knowers can find in nature either necessary production or the lawful necessity of a succession of events . . . . [I]f necessity does indeed exist in a nature understood to be independent of any formative/constitutive power the knower may conceivably possess, then the knower cannot observe, intuit, or otherwise confront it." The best we can hope for in formulating laws of nature is to use them descriptively, detailing how "transitions from one state to another of a physical system — large or small—do in fact take place" (78).

Since the Humean pessimists think that necessity cannot be found in an independent external reality, they attempt to re-interpret the Laplacean model in linguistic-logical terms: "*statements* about the state of a physical science that is regarded as the cause *logically necessitate* statements about the state regarded as the effect" (78). This is currently the dominant view in philosophy of science, and the result is that philosophy is trapped within "a linguistic prison," viewing physical entities not as real things belonging to a real external world but as linguistic postulations belonging to the "ontology" of whatever language the theories about them are expressed in. Despite this major difference in perspective, both factions pursue the traditional "reductionist goal for the unity of science," which requires that all laws aimed at explaining an upper level in a complex hierarchical system be deducible from the laws covering the base level, and that all concepts applying to the upper level be defined in terms of base-level concepts. Even though many of the Humean strain profess to view models of reality as being linguistic constructions, they no less than the others are "dominated by the image of a total (concrete) physical system in continuous progression from state to state under eternal laws that mandate just that progress and no other" (80).

The problems with the scientific doctrine of causality only get worse when you try to apply it to complex physical systems such as human beings, as Pols illustrates in chapter 4, "Mind and the Scientific Doctrine of Causality." Everything that exists and might be studied by science is part of an unimaginably complex universe of nested physical systems, so the Laplacean ideal of

state-to-state research and analysis is in practice supplanted by treating a given entity in relative isolation from the rest of the universe and as a relatively stable structure and attendant substructures, within which some specific thing is happening that we want to understand. However, we don't know how or what to add to the currently understood laws of nature to allow a causal analysis of any relatively complex part of the brain; nor do we know how to establish the initial conditions of such a part of the brain. Thus, rather than treating the whole brain, for instance, as a physical system moving from total state to total state, in practice, scientists instead adopt the more practical cause-effect model which treats one brain event as resulting in another brain event. The problem with this approach, however, is that although we know with certainty that, for instance, a complex pattern of guided electrochemical impulses is essential to vision, we don't know just how those impulses contribute to vision, let alone how they contribute to our rational awareness that we are seeing something.

Beyond this, there is the problem of how complex biological structures arise, both phylogenetically and ontogenetically. Pöls points out the inadequacies of the current neo-Darwinist paradigm in evolutionary theory, which follows standard scientific doctrine in seeking to discredit and eliminate the idea of final causation. Although teleology appears to be the case, neo-Darwinists argue, given enough time all of the plants and animals and all of their characteristics could have arisen by chance operating through natural selection. If there were such a thing as *absolute* chance, their argument would succeed; but the consensus is that there is only *relative* (Laplacean) chance, the kind that "can be eliminated by better knowledge, although such knowledge is sometimes difficult to come by and hardly worth the trouble" (85). Neo-Darwinists indeed do adopt this viewpoint, holding as well the standard view of a smooth, continuous, necessitated movement from state to state. What they do not acknowledge is that relative chance does not have the same teleology-banning implications that absolute chance does, hence they persist in their opposition to purpose and final causation.

Another problem facing the accepted model of causality is the appearance-reality clash that shows up between common sense and the contemporary tendency to try to explain away the causality governing larger structures such as living organisms or the human brain in favor of that operating on the microentities that make them up. While scientists have basically abandoned the attempt to reduce higher-level laws and concepts to lower-level ones, many still try to argue that causality only actually *works* in the entities at the lowest, base level of nature. We should, they say, take microentities more seriously, i.e., as being more real and causally significant, than the organisms they constitute, and we should regard organisms as aggregates rather than integrated wholes. On the other hand, the mind tends to regard at least some macroentities as capable of rather serious things such as responsible action for which there is some causal and explanatory significance. Philosophers who deny the power of the mind to know an independent reality, however, claim that the commonsense idea that large systems and, in particular, the minds that belong

to such systems as ourselves have a causal significance is based on some kind of deception: “. . . *the antireductive disposition of common sense is nothing more than a disposition to take an appearance for a reality*” (89). They blame mind for generating the appearance and then taking it as reality; mind by its very nature reacts to what it wants to know by making something else and then taking that for what it wanted to know in the first place. We take complexes of electrons to be physical objects, we take lights waves to be colors, we take linguistic constructs to be reality. This creates a deep problem for materialism, which views mind as being causally generated by the physical operations of microentities—and also as itself generating the appearance that materialism’s view is not complete—and also that mind too is an appearance. Despite this, the mind must also break free from its being a causally dependent appearance-generating appearance and somehow identify how things really are. Materialism and its view of mind can hardly be defended, when their very premises and conclusions seem to destroy the possibility of any such defense.

To depict starkly the difference between the standard scientific model of causality and the view that living beings are causal hierarchies, Pols proposes a thought experiment that illustrates the pitfall of any attempt to resolve the mind-body question via neurophysiology. First, he says, assume that an omniscient scientist could, at the start of a complex rational act, establish all the relevant conditions operating in the person carrying out the action, without interfering with that action in any way. Second, assume that the scientist knows all of the relevant laws of physics and chemistry and physiology. Third, assume that the scientist doesn’t know anything about what the person being studied is thinking about or intends to do. If, in fact, we are causal hierarchies, the omniscient scientist will be unable, despite all of his knowledge, to predict even the *physical* conditions in the nervous system at the end of the action, because the nervous system’s physical behavior is affected by the apex being’s mental functions. Although the person could not have carried out the action without the causal support of the nervous system, the events involving the neurons are not the entire cause of the person’s mental functions. Hence, the scientist will be unable to offer an adequate *physical* explanation of the state of the person’s nervous system at the end of the action (90-1).

Pols opens the second part of his book with some prefatory remarks relating the laws of nature to the findings of science. He notes that there is no conflict between the view of causality as primarily hierarchical and the actual way in which science has progressed historically. Pols suggests that the laws of nature are “an abstraction from, and a codification of, the ordering power of entities whose ontological status is perhaps more fundamental,” the laws of nature being “derivative from the ontic power of primary beings” in general (88). In other words, Pols says, the laws of nature are “regularities extrapolated to a universality that ranges far beyond their empirical base” and thus are “descriptive rather than ontologically determinative,” detailing the recurrent aspects of entities “whose causal structure is more concretely and more adequately understood in terms of a hierarchy of causes.” The closer the phenomena one examines are to the base of such a hierarchy, the more precise and

deterministic in character are the laws describing such phenomena. On the other hand, without assuming the existence of indeterminism or absolute chance in the universe, the use of statistical laws (rather than deterministic laws) may be unavoidable in describing more complex entities. Pols thus assumes that the reason determinism does not play a major role in human action is not because any kind of acausal "Epicurean swerve" is in operation, but because the apex entity in any given hierarchy (human or otherwise) "can determine what is determinable in the pyramid in which it expresses its self-identity, and that the more complex that self-identity is, the less predictable the outcome." He adds, in the final chapter, that "what is open to determination need not be pervaded by some absolute indeterminacy or chance in order to thus be open." It merely needs to be distinct in some sense from what determines it, in the sense that we at the apex of our pyramids "are in some sense distinguishable" from the biochemical processes occurring in our brains" (95-6, 127). While this does not establish free will in the indeterminist sense many claim is necessary to avoid the hegemony of determinism, it goes a long way toward establishing the relative autonomy of living organisms in general, and human beings in particular.

In chapters 5 and 6, Pols presents his positive thesis: "a rational agent is the apex being of a hierarchy of causes and so a primary being." Here he guides us in "deploying the causality intrinsic to the beings we are in order to remove a doctrinal obstacle to the acknowledgment of that causality"—in other words, in focusing on what mind actually, concretely is—first on "the functions of the minds of the apex beings (primary beings) we are," then on "the unity that expresses itself in these several functions, namely, the causality of the apex being that is refracted in these many functions" (95-6). By "the mind itself," Pols means nothing mysterious or obscure, but rather "the *full concreteness*, the *full actuality*, the *wholeness* of mind, the *lived reality* of mind"—in other words, the embodied mind, "the human being who speaks, argues, chooses, feels, and all the rest . . ." Pols steadfastly refuses to consider the mind as less real and concrete than the central nervous system, and he points out that when the central nervous system is studied "just as a biological entity," it is no less "being considered in abstraction from the full concreteness of mind itself" than is mind when we focus on its functions. Moreover, we are familiar with mind "in a way we cannot be familiar with items of the infrastructure [e.g., electrons, neurons, tissues, etc.] that support its deployment." The method by which Pols proposes to "attend to mind itself and . . . discriminate its functions" is the persistent, unswerving application of something that we all use to some degree or another, what many would call "introspection," and what Pols himself calls "the reflexive turn." This approach, he says, is somewhat like common sense, in that "to bring mind itself into view, you need only . . . focus on rational action, either your own or someone else's." In other words, "our familiarity with mind itself is by way of mind's doings: we must perform one of mind's typical functions in order to be familiar with that function. Our familiarity with mind itself, in short, is reflexive" (11-2, 14).

In the first of three applications of reflexivity in chapter 5, "Mind on Its Own

Functions,” Pols lists and describes the functions of the mind, a task clearly similar to the one Descartes undertook in his *Meditations*, which had “disastrous consequences for our understanding of the human mind.” Thus, Pols says, we must “redo the enterprise of Descartes” (97). Even listing mind’s functions requires reflexive use of at least some of them; and if Pols is right about the functions of the mind being causally structured, then using those functions in making this list is an instance of such causality, both depending on and commanding the infrastructure of the causal hierarchy, without however needing to understand how the infrastructure contributes to either the function or our reflexive awareness of it. Pols replicates and adds to Descartes’s list of mental functions. He adds “knowing,” which Descartes apparently took for granted and so did not notice. Pols omits “consciousness” or “awareness,” as did Descartes, because it is such a pervasive feature of mind that it qualifies all of the other functions; it is not strictly synonymous with “mind,” however, since mind sometimes functions unconsciously. Pols also, with Descartes, omits “action” from the list; while one of the functions on the list may sometimes also qualify as an action, it is because it is in that instance serving as “a function of the whole entity/being—perhaps the most comprehensive of its functions.” Thus, Pols conceives of action as being “expressive of the very unity of a rational animal,” a unity that spreads throughout all the distinct functions that are involved in that action. It is such overlapping of the various functions of mind that reveal unequivocally that mind is involved in a particular situation; see Pols’s discussion of knowing for an excellent illustration of this principle. Pols concludes that the union of functions in human action with the “billionfold multiplicity” of the elements in the infrastructure is so intimate that the standard conception of a mind-body *relation* does not fully capture the embodiment of mind (98-100).

Pols’s second application of reflexivity—at once the most complex, difficult part of his positive thesis and the most crucial to its success—is to the two most vital mental functions, namely knowing and making, which he also refers to as “direct knowing” and “the formative function.” He emphasizes, however, that the primary function under reflexive examination here is direct knowing, and he notes in passing that reflexiveness is so natural and familiar “precisely because it is a capacity of direct knowing.” (This is an important insight, as I will explain in one of my critical comments.) Direct knowing, when applied in everyday contexts, is usually accurate, most errors being correctable by closer attention to the object of concern. Reflexive use of direct knowing, however, is anything but an everyday application of it, and moreover it is important because only by “making direct knowing the object of our attention and at the same time *using* that function” can we nail down what direct knowing can and cannot do; only thus can we “make mind itself aware of its own prerogatives” and shed light on the formative function and the other functions of mind (100-1, 112).

Pols makes an important distinction between primary direct knowing and secondary direct knowing (which he also refers to as primary and secondary “rational awareness”). It is here that Pols most rigorously expounds and

defends his position that reality is independent of, while knowable by, the mind. He says that in both its primary and secondary forms, "direct knowing is a realistic function and does not inevitably form, make, produce, or constitute what it knows." The former is "the knowing of concrete things that are available to us by way of the senses," the process by which "we attend to temporospatial beings (natural or artificial) that fall within our size range." The latter is the direct knowing of such artifacts as doctrines, theories, concepts, propositions, words, and language in general, "as well as mathematical objects, "narratives, poems, and the other imaginative structures produced by the non-literary arts." Such artifacts are not temporospatial, "although most of them are associated with temporospatial entities that are symbols for them," and they are all dependent upon the formative function. Once they are made, however, our awareness by means of secondary direct awareness of such "*entia rationis*" as propositions and theories "is just as direct, immediate, and vivid as primary awareness." Further, the term "secondary" refers not to the degree of reality of what is so known, but to the fact that we focus on such things "against a background that is always available by virtue of primary awareness" (101-2, 112). Pols further points out that the most important cognitive use of secondary rational awareness is in the direct awareness of the theoretical objects that the mind has formed, by means of its formative function, in order to know things like black holes and electrons "that cannot be known directly, usually because they do not fall within the temporal and spatial size range of our sensory modalities." Such things can be known even indirectly, however, only because our rational awareness is able to know directly many concrete temporospatial things. Pols is unabashedly bullish on this matter:

Indirect knowing is one of the glories of human nature, if only because science depends on it: it is the outcome of a complex interplay of secondary rational awareness of theory with the primary rational awareness of commonsense things in which theory begins and in which it is later tentatively confirmed or decisively disconfirmed. (113)

Nonetheless, Pols says, we should not be lured by the vast proliferation of *entia rationis* (conceptual products) into thinking that all knowing is indirect, and that we cannot know reality directly. This is the error into which Descartes led several centuries of philosophers, the "negative philosophical judgment about the powers of the mind"—that radical distortion of the function of mind which Pols is at pains to correct in the second reflexive turn. The key is to realize that the formative function, while essential to both secondary direct knowing and indirect knowing, is a distinct function from direct knowing and does not in any way compromise or negate the mind's ability to engage directly with reality. Pols uses the distinction between primary and secondary awareness to account for the way that the mind moves back and forth between theory/language (the rational) and empirical testing (the experiential) in direct knowing:

[W]e deploy the seamless unity of rational awareness in coping with each of those “poles” of knowledge: we know proposition or theory directly, and we know directly (and do not merely experience) that which is empirically relevant to proposition or theory. (113)

Philosophers typically try to split knowing in two, assigning reason to the conceptual-linguistic sphere and experience to the perceptual-feeling sphere, thus failing to realize that primary direct knowing is not a problematic re-combining of reason and experience, but fundamentally instead “an intimate union of rationality on the one hand and experience/awareness on the other . . . [a union] that both takes place in the knower and completes itself in the thing known.” When philosophers try to break up the integrated function of direct knowing into discrete rational and empirical components (e.g., ideas or impressions, concepts or perceptions), they make the mistake of claiming that these in-the-mind things are what we directly know and that they function as cognitive intermediaries by which we indirectly know, believe in, postulate, or construct concrete things in the real world. Not only does this line of thinking lead to skepticism about the reality-attaining power of mind, but it is also self-defeating in two other ways: (1) you cannot even carry out such an analysis “without immediately reinstating the seamless unity of the function . . . [Y]ou would be drawing on the unity of the function of direct knowing to know both a ‘rational’ item and an ‘empirical’ item . . . [I]n short, you would be depending on the unity of the function to deny its unity,” something Aristotle referred to as Reaffirmation through Denial; and (2) even if you could split knowing into discrete parts, you would then be faced with an infinite regress, i.e., of having to analyze both the ‘rational’ and ‘empirical’ parts into “yet other pairs of rational and empirical components,” and so on (101-4).

Against such pitfalls, Pols urges us to recognize that when we engage in direct knowing, we also actualize a plethora of other functions, including (to name a few) conceiving and perceiving, attention and intention, remembrance and anticipation, which “are real enough yet do not exist in a ‘pure’ form,” because of their necessarily being actualized together, “as part of the integrity of (primary) direct knowing” (101-4, 114). With stern eloquence, he reminds us:

Experiencing (in the restricted sense of perceiving) does not vanish when (in secondary direct knowing) we attend to rational items like ideas and concepts. Rationality (in the restricted sense of conceiving) does not vanish when (in primary direct knowing) we attend to experiential items like patches of color . . . No activity of the mind, no matter how formal, no matter how designed to exclude any reliance whatever on any experiential factor, is without some reliance on all the bodily particularity of some

here-and-now. Conversely, any effort to so isolate the here-and-now as to come upon an experience from which all participation of rationality has been excluded turns out to be merely a misguided exercise in direct knowing . . . The experiential always suffuses the rational and the rational always suffuses the experiential . . . When mind *knows directly* . . . rationality and awareness exist in mutual support: . . . reason *experiences* [and] our experience of what we thus attend to is pervaded by *rationality* . . . You cannot suspend the seamless unity of the act of direct knowing when you undertake analysis, anymore than you can suspend the embodiment of mind when you engage in abstract logical reasoning. You can only do what is so often done in theory of knowledge: fail to notice what you are actually doing. (104, 111)

In his third application of reflexivity, Pols seeks to justify and extend the claims he makes with and for direct knowing. By the very nature of knowing, we cannot step outside of direct knowing in order to provide justification of the conclusions of direct knowing. When direct knowing/rational awareness “completes or actualizes itself in something whose being is independent of the knower,” a “satisfaction . . . takes place within the knower,” which, however, “is wholly taken up with the thing known . . . so much so that the only way to bring out its peculiar character is to call it a satisfaction in knowing the other—a satisfaction in acknowledging the known as what it is” (115-6). The satisfaction, being internal to the knower, is “subjective,” but only in the sense of “of the knowing subject,” not in the sense of “self-enclosure [or] isolation [of the knowing subject] from everything that is not a product of the mind”; and being rational, the satisfaction is impersonal and universal. This fulfillment of the “rational awareness, rational subjectivity, or rational consciousness” of a particular knower is correlated with “objectivity, reality, actuality, or being,” all of which are understood to mean that the things that are directly known are independent “of the function that is satisfied in them.” The function of direct knowing is justified if and when it completes itself in that which is other than itself by acknowledging the other to be what it is. In other words, direct knowing “is a self-justifying function,” not in the sense that it is infallible or “exempt from error in any particular instance,” but that error can exist only because of the possibility of avoiding error. The universal character of one’s rational awareness

transcends each instance it is integral with, and so possesses a general authority that is not touched by its failure in a particular instance. Knowing yourself mistaken about just *what* is before you in some particular instance, you nevertheless know that the misidentified thing is

other than yourself and so independent of your cognitive act. Confident that your failure can only be defined within the framework of a general competence, you find that you are in a position to try again. (116)

In both primary and secondary rational awareness, Pols says, “you cognitively attain the object and enjoy its otherness, but in doing so you allow the object to take possession of your subjectivity/consciousness.” In primary rational awareness, where the object is a temporospatial item, the distinctness of the two entities involved—the knower and the known—includes the fact that there are two distinct temporospatial locations; that distinctness is overcome insofar as “something *over there* is cognitively possessed *right here* in the subjectivity of the knower.” In secondary rational awareness, even though “beings of reason” (products of cognition and imagination) are not always temporospatially distinct from the knower, due to their often having been formed *by* the knower, they are still essentially distinct from the knower due to their being “*formed to be distinct* from the knower.” Each concept, theory, poem, or novel “has an inner integrity that must be respected.” For instance, “as your mind moves through the parts of a theoretical structure to determine their consistency, coherence, and relevance to the matter at hand, their place in the structure has an otherness from yourself as knower that demands rational respect in a way analogous to the demand made by temporospatial objects.” Pols defines the essence of the relation between knower and known as:

the attainment and enjoyment by the knowing subject of the particularity of the known object—that is, a satisfaction on the part of the knower in just *this* known . . . [T]he otherness of the known object is overcome by you as knower, even while the discrete self-identity of the two beings thus brought together is preserved and acknowledged . . . [O]ne component of your satisfaction as knower is your acknowledging, in the very act of taking cognitive possession of the thing known, the utter independence of that object from the function that attains it. But your satisfaction as knower also includes your celebration of the integrity of your own achievement: as knower you have reached out and brought into yourself an awareness of something other than yourself with which you nevertheless acknowledge an underlying affinity. In short, the complex satisfaction is a satisfaction in both the particular being of the object and the successful deployment of the function that attains the object. (117, 118)

By thus having explicated the cognitive relation between knower and known, Pols says we are now in position to reject the modern view that universals originate in the mind, which uses them to project an *apparent* entity; that what our mind (using universals) presents as being unified, stable entities are actually just “a multiplicity of particular stimuli.” The mind is only able to attribute universals to particulars because “the formative function of mind is also engaged in the transaction,” breaking up the entity’s unity or universality into a number of “particular” universals. While rationally aware of a particular, “we are also aware that it shares in a unity that all particulars share in”: it does not participate in some particular Platonic form, but “in a unity/universality to which the formative function of the mind has responded by producing multiple (particular) Platonic forms” (118-9).

The *coup de grace* to “the negative philosophical judgment about the powers of the human mind” and the “received scientific doctrine of causality” is administered in chapter 6, “Mind at the Apex of a Hierarchy of Causes.” Pols takes us through the fourth and final movement of the reflexive turn by considering the rational actions that utilize mental functions and the agent of those actions, “a human being that acts—the being in which the action originates, out of which the action comes . . . [which] may properly be said to cause the action and so provide an explanation for the existence of the action” (120). Human beings do not, however, cause their acts in the same sense in which the prior movement of one physical object causes the subsequent movement of another; if causality is understood strictly in terms of temporal sequence, the relation between human beings and their acts cannot be instances of causality. We continue to shape and guide our acts, rather than simply initiating them and then having no further causal influence; we have the complex effect we are aiming for in mind, and having it in mind has an influence on the effect coming about: “[T]he telos is effective throughout the sequence of which it is the completion” (122). From this, it is clear that the causality of human action cannot be made intelligible without a considerably broader and more multifaceted model of causality than the one proffered by mainstream science.

The way Pols proposes to transcend the overly narrow scientific model is to consider the causality that the mind of a human being both knows and exercises as it gains direct knowledge of the world in which it lives. He has already shown that direct knowing is capable of grasping the reality of something—whether concrete, perceivable things in the world or things ideas and theories—distinct from and independent of the knower. The things so grasped are both particular and possessed of a kind of ordered unity that is present in everything that exists, however simple or complex, including the mind itself. Pols calls this the “U-factor” (“U” for “universal”). When the formative function produces ideas, theories, etc., those products are then *formed* realities that secondary direct knowing can grasp. By attending to mind itself, we thus find that the cognitive achievements of direct knowing and the creative achievements of the formative function are *causal* achievements as well—and that it is the *mind itself* (as embodied in the infrastructure of the human body) that achieved them. In so doing, we realize that “to be capable of knowing an inde-

pendent reality is to be capable of *being a cause* in a sense that illuminates that achievement.” When we cognitively grasp real things that are other than our particular selves, we are causally responsible for that grasp, as well as for our cognitive grasp of the causality of those other real things, some of whom have the same kind of causality we used in our cognitively grasping them. This is not a mere inference or postulation based on evidence, but the result of our using, as rational agents, a function whose power to grasp reality includes the reflexive grasp of the status of ourselves and others as real entities.

Now that we have established the power of the mind to know the components of its infrastructure, to understand how we rational beings have causal power, and to understand how components of our mind’s infrastructure have causal power, we can see how our minds are situated at the apex of a causal hierarchy. Having used our mental functions in determining their own nature and validity, we can further confidently and validly use them to grasp the fact that those functions are causally dependent on our being the apex of “a causal hierarchy . . . an infrastructure that defines the embodied state of the human mind” (125). Pols uses the metaphor of a pyramid to illustrate how each of us is the apex of a multilevel structure “of causes made up of untold myriads of entities/beings, each of which is the apex of a smaller pyramid of causality. As you deploy your various functions in an act, . . . you also deploy your causality—your power of determining something—down through the multiplicity of the pyramid” (126). Each of us is only *one*, while each level below us is composed of *many* items. While each of these items, by virtue of its own determining power, contributes that power “upwards” to the causality we exercise in rational action, we in turn exercise our own determining power “downwards” over each of those items. The effects we thus produce in the items lower in our pyramids are not the result of a physical process operating in a cause-effect sequence. We cannot activate a mental function without the simultaneous, nonsequential pattern of firing that allows the function to take place. This same pattern pervades the way in which smaller causal pyramids within us determine “downwards” the activities and outcomes of their components that contribute to the pyramids being just those particular pyramids, while their components contribute causally “upwards” to the pyramid’s exercise of determining power. In the same way that our self-identities are dependent on the particularity of our own pyramids, so in general are the items within our own pyramids related to the items within their pyramids.

Further, in referring to living organisms as entities or beings that are pyramids of entities or beings, Pols is signaling his disagreement with the current fashion of regarding “functional levels” as being more respectable than the things that possess those levels of functioning and whose carrying out those functions raise the issue of what a function is. His ontology regards entities as the fundamental kinds of things in the world; and although we commonly refer to “anything that we can single out by its apparent unity from the rest of the enviroing world” as an entity, the kinds of entities that carry out actions, he says, are the *primary* entities (127). For this reason, he refers to primary entities as exercising “ontic causality” or “ontic power;” and he adds that, to the

extent that the entities at lower levels of our organismic pyramids carry out act-like functions, they too can be said to exercise ontic causality over their own components and to function, in a more limited sense, as primary beings. Ontic causality is universal, existing in and transcending every individual, unified thing that exists; it is the U-factor mentioned earlier. This is Pols's ultimate reply to microentity reductionism:

[W]e have dismissed the claim that the (transcendent) nature of things has its *locus operandi* only in the microentities of the base level. For that we have substituted the claim that its *locus operandi* is in the apex of each primary being from the most evanescent particle to such highly complex beings as Newton and Mozart. (132)

Here, at last, Pols reveals the full structure of his model of causality. The power we exercise, on any given level of our organismic pyramids, in any of our functions or actions is *temporal*, taking time to occur or be carried out, and in a sense "*horizontal*", happening between distinguishable entities in a cause-effect manner: "a temporal sequence in which two distinct items can be discriminated— one in which the power originates (you or me), the other on which the power is exercised (some item in the world around us)" (131). We affect other entities in the world and cause things to happen in the world; parts of our bodies affect other parts of our bodies and cause things to happen inside us. This physical mode of causation (which many think is the only kind there is) Pols refers to as "transeunt" causality, in contrast with what the medievals called "immanent" causality, and which Pols refers to as "ontic" causality: the "*vertical*" and *atemporal* causal relationship *between levels of an entity*. The upward and downward causality that we and our body parts exercise is nontemporal, in the sense that in exercising it, we do not do anything somewhere else "whose impact or influence in the multiplicity of the level below [us] *only* appears there *after* [we] deploy it" (129).

You do not think and afterwards produce electrical patterns to which your thinking contributes. So also with the support given your act by the neuronal level: each neuron does not do something whose impact or influence only appears afterwards in your thought. (129)

In contrast with the distinct entities involved in transeunt causality, the relationship between interacting levels of a hierarchy is ambiguous. In one sense, each of us at the apex of our pyramid is "identical with the multiplicity of functioning items" in our pyramid; in another respect, that self-identity is asymmetrical, in that the apex is a One and its functional items are a Many:

The self-identical actuality of each being that possesses immanent, or ontic, causality thus consists in an asymmetrical union of a universal One with the particularity of a Many on which it confers unity. Any being that is the apex being of a pyramid of many beings is a *particular* being by virtue of that pyramid, but it is *one being* by virtue of a One that is not unqualifiedly particular to it . . . [W]e are not considering a radical plurality of Ones but a plurality of Ones that share in a universal (thus transcendent) One . . . It is the Many intrinsic to any such being which makes it particular; it is the One intrinsic to it which makes it not merely a particular. We are such beings in all our occasions but most vividly so in the exercise of our mental functions. (133)

Further, says Pols, the union of the apex and lower levels of the pyramid is so intimate that the term “relation” is not adequate to describe how they are . . . related! The pitfall comes in regarding mind and body in a Cartesian dualistic manner as being two functional levels—consciousness and neuronal—each of which is “ontologically complete in itself [and] capable of acting on the other” (129, 130). Consciousness is not something that has ontic causal power that it exercises over neurons, and vice versa. It is we as unified entities that *achieve* consciousness of things in the world by exercising our ontic causality over neurons and receiving support from neurons. Consciousness, that is, is not a *source* of causal power, but an *outcome* of it; it results when we take action in the world. It is not consciousness per se, but we, *as conscious beings* that cause things to happen.

The hierarchical model of causality also allows a clearer understanding of the nature of the relation between knower and known. Here Pols is careful to draw the vital distinction between the causal relation between the known and the knower and the cognitive relation between the knower and the known. If rational awareness really does reach out and actualize itself in something other than itself, then the temporal, sequential, cause-effect, transeunt causal relations of the scientific doctrine of causality are only part of the explanation of the relation between an act of direct knowing and an entity or situation toward which it is directed. This basic cognitive achievement also essentially consists in the nontemporal, immanent, ontic causality exercised both by the apex being (knower) and by all parts of its pyramid. Rational awareness begins with an act of attention by the knower. The apex and infrastructure are nontemporally, ontically united, as the knower engages in the temporal, transeunt act of rational awareness of the known, and is otherwise temporally, transeuntly linked to the thing or situation known. In particular, the known has a causal affect on the knower’s nervous system, which in turn affects the ontic support given by the nervous system to the knower. The result of this complicated causal pattern is that the apex being cognitively responds not to its own infra-

structure as affected by the known, but to the known itself. This is how “*the knower achieves rational awareness of the known*” (134). The knower cannot extract an independent reality out of the knower’s own infrastructure, however; at most, what the knower possesses in its infrastructure is some sort of neural *representation*, mapping, or coding, as against the known, which is the *original*. It may be that the knower uses its internal representation of the known in reacting to the known, but it is the *known* that is *known*, not its representations in the knower. Rational awareness thus requires more than just the transeunt causal relation between the known and the knower’s nervous system and the immanent causal relation between the knower’s nervous system and the knower. Pols says that the additional necessary factor is the “affinity” between knower and known that derives from a universal factor, i.e., “an ordering power they share in,” though he does not think we are capable of know exactly how it works, just that it *does*.

[O]nly mind itself, in each of us, can determine what the functions of our mind can and cannot do. And since even those of us who are materialists constantly make determinations that this or that is *truly the case*, it seems that all of us acknowledge, at some level of discourse, a reality-attaining competence that belongs to mind itself. (135)

Thus does Pols sweep aside once and for all the negative philosophical judgment about the power of mind, which “led to the dismal conviction that we can know neither other beings nor ourselves directly and so cannot know any causal significance they have” (137). And he does so with such elegant consideration for the reader. He graciously concedes that his own positive thesis is not exempt from the fact that “doctrines can stand in the way of actuality,” and he generously invites the reader to apply what is seen by a “focus on what our minds actually do” in correcting “what is inadequate in what [he has] written” (20). No double standard here! Although hard pressed to find major fault with anything Pols has said, I will, however, offer these observations on what are some relatively minor points.

Although Pols studiously avoids the term “introspection,” in favor of phrases such as “mind attending to itself” or “the inwardness of mind” or “the reflexive turn,” it is clear that he regards such familiarity with mind, however labeled, as a real process, giving direct awareness of another real process. Yet, he makes one statement that I regard as incorrect, followed closely by another that I regard as uncharacteristically and unduly pessimistic. The latter first: “We may be unable to give an adequate account of what reflexiveness is, and probably we shall never be in a position to say how it is possible—how, for instance, some infrastructure items might subserve reflexiveness” (14). Surely we already know, in a general sort of way, how the direct awareness of reflexivity or introspection most likely takes place. In regard to two forms of direct sensory awareness—externally directed perception (sometimes called “exteroception”) and internally directed perception of bodily conditions (“sensation”

or “interoception”)—it has long been known that the brain receives and interprets signals transmitted from receptor organs (i.e., certain tissues sensitive to patterned energy associated with the object of awareness). There is no apparent reason why mind in its reflexive mode is any the less in need of a physical infrastructure than is sensory awareness, or mind otherwise for that matter. It is thus theoretically parsimonious and hardly a stretch of the imagination to suppose that, in various parts of the brain, numerous groups of tissue, however presently obscure and difficult to detect in their functioning, are similarly capable of responding to patterned energy associated with the brain’s own mental functions—those groups of tissue thus serving opportunistically as an itinerant receptor organ attuned to mental processes carried out in neighboring tissue regions, and thereby functioning as the infrastructure of introspection. This seemingly reasonable speculation is, I submit, more in the overall optimistic spirit of Pols’s approach to the mind-body problem and philosophy in general—and in no way negates his important point that reflexiveness, like all mind’s operations, is only *incompletely* explained by a temporally causal, infrastructure explanation.

As for the first statement, “We do not experience mind itself as we experience (by virtue of mind itself) a color, a smell, or an ordinary physical object like a chair,” there is a respect in which Pols is correct: we experience colors, smells, etc. by using the senses to attend to physical reality, not to the senses themselves, whereas we experience mind by using mind to attend to itself. However, the general kind of awareness is the same in each case; both sense perception and introspection are forms of direct awareness—moreover (if I am correct), using tissues (whether peripheral or internal to the brain itself) that function as the infrastructure of *direct* awareness (viz., as receptor organs for collecting patterned energy and relaying the collected data to some region of the brain for further processing). Furthermore, just as the mind as we are reflexively aware of it is no less real than its physical infrastructure of which we are not (and need not be) directly aware but which supports and enables its functioning, so too are the colors, odors, etc. of physical entities as we are perceptually aware of them no less real than the physical infrastructure of those entities (viz., the atomic and molecular structures) that support and enable the interaction of their colors and odors (via streams of patterned light and airborne chemicals) with our sensory organs.

Another salient virtue of Pols’s thinking and writing is the clarity that results from his conscientious and careful analysis and use of terminology. Whenever there is the least chance for misunderstanding, due to conflicting historical usages of a term, Pols guides the reader through the maze of terminology shifts and follows through when appropriate by joining with a forward slash two or more terms taken as synonymous or necessarily related in that instance. Due to the degree of complexity and amount of potential confusion in the topics linked to the mind-body problem, there is thus a proliferation of items such as “states/events,” “form/nature/essence,” “ideas/concepts,” “Forms/Ideas,” “entity/being,” “subject/substance.” Far from seeing this as a flaw, I regard it as a methodological or stylistic virtue. What I would like to have

seen, however, is an even more exhaustive application of the technique. With all the care Pols took to clarify the concept of an entity's essence or nature, for instance, it struck me as odd that he did not also incorporate the oft-encountered synonym "identity" into that discussion. Similarly, although Pols clearly distinguished between "a being" in the sense of an individual entity and "being" in the sense of existence, there was no conjoining of the latter terms into "existence/being" that I could detect. Another example, perhaps more helpful, would be the acknowledgment of the term "introspection" and, for reasons noted in the previous paragraph, its conjoining into "reflexivity/introspection." As I said, however, these are relative minor quibbles, far outweighed by the considerable good that Pols accomplishes with his illuminating treatment of "essence," "substance," and the "object(ive)," "subject(ive)" pair. The latter discussion, revealing the flip-flop in meanings perpetrated by the Cartesian-Kantian paradigm, ranks in insightfulness and clarity with that in John Deely's *Basics of Semiotics* (Bloomington, Indiana: Indiana University Press, 1990).

The one other comment I must make is that, despite the laudably clear-headed way in which Pols approaches his subject, some of the related issues he covers—especially the historical development of the "received doctrine of causality"—are very *difficult*. In recent years, for instance, I have witnessed the foundering of numerous mind-body and free will debates by thinkers otherwise very Aristotelian, because they were not able to sufficiently break out of the Humean model of event-causation to realize that causality characterizing mind is not fundamentally a matter of antecedent conditions. For this reason, I am not confident that Pols's negative case will have the impact that it needs to have, if the narrow, inaccurate Humean perspective is to be supplanted by the Aristotelian entity-action concept of causality, and its corollary, the Platonic-Aristotelian apex-infrastructure model of the causality of primary entities.

On the other hand, that is all the more reason to appreciate the skill with which Pols marshals his positive arguments. He has crafted a major advance in philosophical methodology—a carefully formulated way of thinking (reflexivity/introspection) that is as challenging and revelatory as the Socratic method—and has applied it in a way that establishes the reality and potency of the mind, while lifting up the dedicated reader's ability to exercise that very potency. He has also formulated a multifaceted model of causality that does full justice to the subject, allowing thinkers to grasp not only the nature of the physical world but also the nature of the living beings that inhabit it and, (some of them) try to understand it. Had a thinker of Pols's caliber been active 400 years ago, we might have been spared the long detour of modern post-Cartesian philosophy, and all of its dreadful consequences, intellectual and social. Even now, however, for the world to enjoy the full salutary benefits of Pols's way of thinking, it will take more than just a book review here or there by an enthused reader to do the trick. His ideas and methods must be taught and spread in university philosophy departments where, if all goes well, they will eventually trickle down to the theorists in the sciences and humanities. If any of the readers of this review are inspired to help begin such a trickle-down process, it will have achieved its purpose.

## Book Reviews:

Hartle, Ann. *Self-Knowledge in the Age of Theory*.  
Lanham, Maryland and London: Rowman & Littlefield  
Publishers, Inc., 1997.

The deference shown in Aristotle's writings toward the *endoxa*, the "respectable opinions" of the city, is well known. Aristotle's deeds in no way contradicted his deferential words when he fled his adopted city on charges of atheism, in order, so tradition has it, to save Athens from sinning against philosophy a second time. For although the philosopher will normally suppose the city's views to be truthful, this is not to assume that Athens, or any other human community, always knows its truths fully for what they are. The difficulty made manifest by the Peripatetic's flight is that philosophic interrogation of received opinion cannot help but seem suspect to the city's many non-philosophers: do not such inquiries betray the philosopher's hidden ambitions to rule, or worse, his disdain for all that the city admires? What such citizens fail to see is that the philosopher is actually so impoverished, so uncertain of his possessions, that he can only come into his own by constant comparison of his estate with the doxic images of that which he himself proves, but only in retrospect and quite apart from his doing, to be the original. The experience of ancient philosophy would seem therefore to show that a degree of civic hostility toward philosophy is the understandable, if also defective, human norm. To pay his respects to respectable opinion, the philosopher must run the risk of appearing disrespectful.

Modernity seems, however, to have arranged matters quite differently. "Public opinion" is nowadays remarkably accommodating to philosophy, or at least, to something that bears more than a passing resemblance to philosophy. In the words of Walker Percy, ours is "the age of theory." As regards most anything worth talking about, ordinary citizens now readily defer to the experts, who generously offer their teachings to all and sundry. When a modern academic perishes, it is never from hemlock; it is from failure to publish; and while some might consider tenure at a third-rate state institution a kind of death, this grim fate invariably includes a benefits package considerably more expansive than that offered by the Prytaneum. To be sure, one might very well doubt whether the public embrace of "theory" constitutes an unqualified advancement in learning over Greek intolerance. Percy himself explores the results of theory's largesse to great comic effect in his *Lost in the Cosmos: the Last Self-Help Book* (New York and London: Washington Square Press, 1984). In *Self-Knowledge in the Age of Theory*, Ann Hartle takes up the same problem in a more serious or academic way. However, hers is not merely, or even primarily, a "negative" or "critical" work. In the spirit of premodern philosophy, she seeks to provide an account of the nature of self-knowledge precisely through a consideration of three failed attempts to do the same. Not theory, but philosophy, is for her the proper path to genuine self-knowledge, and she attempts to dis-

play the distinction between the two through an examination of three theoretical pretenders to philosophy's vacant throne.

Such indirection might seem to be, but is not either *faute de mieux*, or sour grapes. For Hartle as for Socrates, the life of philosophy is strangely dependent upon sub-philosophical knowledge, and its attendant ignorance (xiii), and so culminates in the still stranger conviction that philosophic knowledge is indistinguishable from philosophic ignorance (146). It is surely no fault of hers that her quarrel with modern "theory" does not, in Socratic manner, threaten to roil the waters of public life, and thereby call attention to her activity. Still less is she to be blamed for the unprepossessing character of the theories she treats, which rarely evince the charms of the old endoxa they were meant to replace. It seems that once the people, or the age, has been persuaded to proclaim theory king, it does not much to matter to us what particular theory claims title to rule. To exaggerate only slightly, "theories" are to modern rule what persons once were to constitutional office: King Theory is dead, long live King Theory! Indeed, to put philosophical questions to a ruling theory will doubtless be construed by some as part of the rites of theoretical succession. We moderns are both more attached to theory, and less attached to theories, than were the Athenians to their endoxa.

Strictly speaking, "theory" is the target of only the first of this book's five chapters, which gathers under that heading a variety of contemporary teachings whose common concern may be said to be the nature of human intelligence. No mention need here be made of the complete cast of characters. Let E. O. Wilson and W. V. Quine—both Harvard men, as it happens—stand for the whole company. What unites this disparate body? According to Hartle, it is the conviction that all serious thinking is theorizing, and that all theorizing is the manipulation of a web of univocal meanings (a "theory") abstracted from, and so purified of, the endless ambiguities present in what is deemed, from the theorist's point of view, to be "pre-theoretical experience" (1–13). Hartle shows the "insufficiency" of theory so conceived by invoking the argument by retorsion. Thus, theory fails theoretically so to speak because its very terms make it impossible to maintain the distinction between "appearance" and "reality," despite the fact that the distinction is indispensable to the theorist's claim to be in the know (13–18). Modern theory also undercuts the terms of its practice, insofar as it at once assumes the distinctiveness of human "agency and autonomy" and denies it (18–25). In the end, therefore, theory has little of real interest to say either about the world's self-disclosure to us in speech, or about our self-disclosure in deed to the world.

The second chapter proposes to treat "anti-theory." The figures Hartle selects to represent this "mode of thought" (27) are, not unexpectedly, men like Richard Rorty, Stanley Fish, Michel Foucault, and Jacques Derrida. Hartle's concerns are not primarily genealogical, so she here passes up the opportunity to explore telling differences between the life of anti-theory in Europe and in America. In any case, the neologism "anti-theory," like the by now badly shopworn term "post-modern," identifies a position that defines itself by reaction or negation. Because Hartle is not herself reactive she manages to portray the motives of "anti-theory" with some sympathy; the limitations the anti-theorists detect in their opponents are, as she has already demonstrated, real enough.

Alas, we tend to become parodies of what we hate, and as we see from her succinct summary of anti-theory, Rorty and Foucault succumb to the same sort of problem as do the lovers of theory: anti-theorists fail the test of self-reference. As in the preceding chapter, then, Hartle examines what her interlocutors mean by meaning (29–34) in order to establish their inability to speak meaningfully of human speaking (34–42) and human doing (42–50).

The third chapter of *Self-Knowledge in the Age of Theory* takes up what most of us would surely take to be a more congenial stance, which Hartle identifies by the name of "narrative." The abstruse talk of a Quine or a Rorty is surely an acquired taste, but who can resist the charm of a good story? Of course, there is no such thing as a narrative tout court; there are only narratives. And here we confront an obvious difficulty. Aboriginal mythologizing, *The Book of the Dead*, the *Illiad*, Herodotus's *Histories*, Plato's *Phaedo*, the Book of Job, the Gospel of St. Mark, Augustine's *Confessions*, *The Consolation of Philosophy*, the *Inferno*, *The History of Troilus and Cressida*, *The Voyage of the Beagle*, and television talk-show interviews all involve "narrative" in some sense; but that which these various "narratives" narrate serves more to distinguish them than some common means of expression serves to unite them. Out of decency, perhaps, Hartle does not refer to learned discussion of this "form of self-understanding" (51) as "narrative-theory" (but compare 73); nevertheless, we still clearly have to do with yet another theoretical mode.

By way of ordering our thoughts about narrative as such, Hartle offers the precision of three distinct "forms" of narrative: the "defining myth or story . . . of a community," "literature," and "self-narrative" (54–55). In all instances "narratives" draw close to the old *endoxa*, preserving human particularity from the predations of "theory" and "anti-theory," while still allowing some space for the detachment of rational reflection, at least in principle, or on occasion. Alisdair MacIntyre is, accordingly, the greatest prominence considered in the chapter. Hartle advances a judicious appraisal of attempts made by him, and others, to move beyond the limits imposed by modern rationalism upon human self-understanding. She argues quite convincingly that even or especially apart from its content the narrative form does seem to offer distinct advantages in accounting for the place of appearances of human life, and the place of human initiative within the world; but as she also demonstrates, the strengths of narrative are also its weaknesses. Narrative rescues something of human particularity from theory's aloofness and anti-theory's willfulness, but it, too fails to do justice to the desire to know oneself as this particular human being (75–83). In the concluding two chapters of *Self-Knowledge in an Age of Theory*, Hartle offers a more direct account of philosophy as a response to the discovery of human "interiority," by which she means the particular way in which the human being is realized as a particular instance of a particular kind. In the fourth chapter she advances the bold claim that "ancient philosophy" does not confront interiority in all its radicalness (107; cf. 86–88, 90–91): despite the best efforts of Aristotle, for example, to accord a distinctive place in his philosophy to human distinctiveness, he seems unable to do perfect justice to the strange incongruities obtaining between the individual human part and the natural whole. Accordingly, Hartle turns to Augustine, and to his greatest modern student, Pascal in order to extract a provisional description of the "self" that is the

theme of her book, although in an equally bold claim, she asserts that Montaigne is the single most accomplished exegete of human interiority known to us, a claim she promises to make good on at some later date (xv-xvi, 83, 146, 179). This leads, in the final chapter, to some helpful hints as to the nature of philosophy's overcoming of the insufficiencies of theory, anti-theory, and narrative.

By necessity, this summary description of the book's trajectory overlooks the many valuable observations that inform it along the way, for example its retrieval of "rhetoric" from the reductionism of "power politics" (40-50). I conclude these remarks with a general comment, however.

It is not self-evident why the specific "theories" Hartle evaluates in this thoroughly engaging book, for all their influence over the age, have come to assume such authority over modern self-understanding. Of course, to the extent that self-knowledge is never something given, but always to be earned (and that in full awareness that it can never be earned completely), any theory will prove to have its allure. Still, it is, again, rather striking that two of the three theoretical stances she ponders fall considerably short of the human appeal of the *endoxa* they helped to displace; as for the third, it is oddly unaware of the resistance of the *endoxa* it hopes to recover to the necessity of "theory" in an older sense of that word. I am not suggesting that Hartle is to be blamed for the homeliness of her theoretical opponents. To the contrary, she is much to be commended for her dialectical ability not only to see for herself but also to show others that these would-be emperors are, if not naked, than more scantily clad than their conduct would suggest. And yet as she herself concludes, philosophical self-knowledge "is no more, and no less, than the knowledge of ignorance" (146), which suggests that in an age of theory, self-knowledge demands extensive reconsideration of the properly *philosophical* knowledge that first gave rise to modernity's theoretical ignorance. In view of her own claim that philosophical self-knowledge "is sought in the conversation of friends, a conversation not to be distinguished from the philosophical engagement with the history of philosophy" (143), we are surely entitled to conclude that *Self-Knowledge in the Age of Theory*, notwithstanding its many virtues, means in fact merely to be a protreptic to a more intensive confrontation with its theme. I, for one, shall await her promised study of Montaigne with some eagerness.

John C. McCarthy

## Book Reviews:

*The Significance of Free Will*  
by Robert Kane. New York:  
Oxford University Press, 1996.

In skimming through the notes and references to Robert Kane's critical overview of the last few decades of literature on the free will debate, one cannot help but be awed by the sheer volume of work that has been done on the subject — as well as its persistence as one of the great problems of philosophy. (Millennialist readers no doubt nod approvingly as they read Kane's citation of 12th century Persian poet Jalalu'ddin Rumi, who opined that the free will "disputation" would continue to the Judgment Day.) More importantly, in following Kane as he develops his own challenging incompatibilist position during the course of this survey, one cannot help but admire the various methodological strengths of his approach, including especially the ability to see and elucidate a simple, yet powerful analytical structure on which to hang all of the intricate twists and turns of the fabric of the debate — although, as I hope to show shortly, it seems that he has not laid out that structure in quite the way that he has labeled it.

Kane is very sensitive to the need to come up with not only a coherent model of free will, but also one that corresponds to reality — in other words, a model that not only meshes in a non-contradictory way with other basic ideas and values (viz., the nature of causality and one's needs as a human being), but also refers to something real and intelligible. (Readers of this journal may be struck, as this reviewer was, by the similarity of Kane's approach to the "integration-reduction" model espoused by Leonard Peikoff in *Objectivism: the Philosophy of Ayn Rand*, New York: Dutton, 1991, the task of gaining valid knowledge being taken to be twofold: derive with logic from one's perceptually given data a system of abstract ideas that cohere with one another, and make sure those ideas adhere to reality by being able to logically trace their connection back to a foundation in perceptual data.) The task, Kane says, is like climbing a mountain (the "ascent" problem) and getting back down again (the "descent" problem):

[A]bstract arguments for incompatibilism that seem to get us to the top of the mountain are not good enough if we can't get down the other side by making intelligible the incompatibilist freedom these arguments require. The air is cold and thin up there on Incompatibilist Mountain, and if one stays up there for any length of time without getting down the other side, one's mind becomes clouded in mist and is visited by visions of noumenal selves, nonoccurrent causes, transempirical egos, and other fantasies. (14)

One also has to appreciate Kane's unwillingness to rest with the traditional impasses and the attendant dismissal of further discussion, his pioneering urge to "dig more deeply into the conflicting intuitions that lie beneath the disagreements," to delve into "whole passages in the labyrinth of free will as yet unexplored" (5, 16). Kane gratefully accepts the compatibilist acknowledgment of various kinds of free will "worth wanting," which amount to *acting* freely, i.e., being "unhindered in the pursuit of your purposes." He pushes on, however, to consider the kind of freedom that is not compatible with even soft determinism, namely, "the power of agents to be the ultimate creators (or originators) and sustainers of their own ends or purposes" (4). Is this kind of free will compatible with determinism? Why do we, or should we, want it? Does such a freedom make sense? Does it actually exist? In dealing with these four questions — the questions of Compatibility and Significance (grouped together by Kane as the "ascent problem") and the questions of Intelligibility and Existence (treated under the "descent problem"), respectively — he seeks to refute the compatibilist claims that free will as "ultimate creation of purposes" should be dismissed as obscure and unintelligible and as not fitting "the modern image of human beings in the natural and social sciences" (5).

As said above, Kane seems to have erred in grouping and/or labeling these problems. It appears to this reviewer that Compatibility and Significance, while naturally (as Kane notes) being considered together, comprise an ascent and descent movement, as do Intelligibility and Existence. In the first movement, we consider the *plausibility* and *real value* of incompatibilist free will, the possibility of a kind of non-determinist free will that is "worth wanting." In the second movement, we consider the *intelligibility* and *real existence* of incompatibilist free will, the possibility of a kind of non-determinist free will that has a place "in the natural order where we exist and exercise our freedom" (184). Plausibility and intelligibility both (not just the former) pertain to ascent, to focusing on the coherence and non-contradictory status of ideas, while real value and real existence pertain to descent, to focusing on the adherence of ideas to reality. In support of this claim, I note that Kane states *all* of his "theses" — sentences or paragraphs elaborating his theoretical view of incompatibilist free will — in the sections on Compatibility and Intelligibility, *ninety* of them in all, and none in the sections on Significance and Existence. Secondly, I note that the sections on Compatibility and Intelligibility contain a total of 122 pages of material, while the sections on Significance (for which the book is named!) and Existence together amount to a scant 34 pages.

So, not once but twice, it appears, Kane takes us up "Incompatibilist Mountain" in a noticeably lengthier and more laborious ascent, then back down in a breathtakingly swift, brief descent. There is nothing wrong with climbing the mountain twice, so long as one is clear about what one is doing; and it would probably not be too difficult to arrive at a suitable relabelling of the sections comprising those movements. It may well be, however, that Kane's Significance chapter — and thus the very need and justification of his book! — is fatally compromised by his having hung its conclusion on the (as I will argue) at least somewhat suspect notion of "objective worth," which in turn may be a result of his failure to realize that Significance is an issue of descent, viz., of correspondence to reality of one's values. Despite this methodological problem,

though, does Kane nonetheless succeed in meeting the compatibilist challenge(s)? Not entirely, as I hope to show in the remainder of this review, which will briefly consider and evaluate Kane's interesting and sometimes innovative answers to these problems and questions.

In addressing the Compatibility Question, Kane says that while freedom from constraint, coercion, and compulsion are all worthwhile as well as compatible with determinism, another important and more basic kind of freedom, the "power of agents to be ultimate creators and sustainers of their own ends and purposes," is not (32). Although two frequently asserted requirements for the latter freedom are often regarded as equivalent, Kane believes that clearly distinguishing them is a necessary condition for moving beyond the current impasse between compatibilists and incompatibilists. One of these requirements is that an agent "could have done otherwise," that the agent has "Alternate Possibilities" (AP) — the other that the agent of an action is the source or explanation of its action, that the agent has "Ultimate Responsibility" (UR). Unlike many other incompatibilists, Kane holds that the case for free will requires not only AP but also UR, and that AP is only seen to be necessary for free will by invoking UR. Indeed, Kane says, the basic reason for the stalemate on the question of whether someone "could have done otherwise" is that compatibilists do not take the UR condition seriously. Thus, the two sides unwittingly argue at cross purposes. Were they to explicitly acknowledge the necessity of UR, it would be clearly seen that free will, "freedom to do otherwise," cannot be accounted for in compatibilist terms.

Kane unpacks the two facets of UR as follows:

[I]n order to be ultimately responsible for being what you are (or for having the character and motives you do have) there must have been something you could have voluntarily done (or omitted) at some time or other that would have made a difference in what you are (or in the character and motives you now have) . . . . [I]f antecedent conditions and laws of nature (or prior character and motives) provide a sufficient reason or explanation for an action, then . . . the agent must be responsible for at least some of those explaining conditions. Something the agent willingly did or omitted must have made a difference in whether or not these explaining conditions were the case. (72, 73)

By "sufficient reason," Kane means either a logically sufficient condition, a sufficient cause (i.e., an antecedent condition governed by a law of nature), or a sufficient motive. By the latter, Kane means to include the requirement that ultimately responsible agents be the source of not only their actions but also their *will* to engage in those actions. In order to avoid an infinite regress of voluntary actions for which one is personally responsible, he argues, there must

be some such voluntary actions (“self-forming actions”) that were not determined (i.e., had no sufficient causes). What one must have been able to voluntarily do that would make a difference in whether or not voluntary actions occurred “is simply *doing otherwise*, rather than doing something else that would have causally contributed to their not occurring” (75).

Thus, Kane says, “could have done otherwise” has a legitimate meaning beyond the thinner sense attached to it by compatibilists. Moreover, one can also be ultimately responsible for actions when one could *not* have done otherwise, so long as those actions were willed actions (issuing from one’s character and motives) and one was responsible by earlier self-forming actions for the character and motives from which one’s actions followed. Acts “of one’s own free will” can be *determined* by one’s will and still be the *product* of one’s free will. In regard to the *forming* of that will, however, determinism cannot be the sole factor, if one is to be held responsible for its formation — not to mention the actions flowing from its exercise. The problem I see with this position is that it does not allow for the possibility that the attainment or formation of one’s free will is not a voluntary achievement but instead an *emergent capability* that arises as a matter of course in the (relatively) healthy development of a (relatively) normal human being. One at some point may simply be *provided* with this capability by the unfolding of one’s genetic endowment, much as one is provided — after sufficient development and absent excessive constraint, coercion, compulsion, or oppression — with one’s sexuality or one’s capacity to realize that objects continue to exist when unseen. What one *then* does with this capacity, is what is “up to one,” that for which one can be held accountable; but even how one’s will becomes “set one way,” a way in which one most wants to act or is *intending* or *trying* to act, may be determined before one has anything to say about it voluntarily (114). I hasten to add that I am not asserting that this is true and that indeterministic free will is false. Indeed, I don’t think we yet know what is the case. More information is needed: the jury is still out as to whether self-forming willings are voluntary acts or pre-programmed developmental events.

Kane, however, on the strength of his confidence that he has demonstrated the plausibility of free will cum ultimate responsibility, next considers the Significance Question: “What is so important about ultimate responsibility that should make a freedom requiring it ‘worth wanting’?” Many of us, after all, believe we want “sole authorship” or “underived origination” over our actions. Are we realistic in wanting this? Traditionally, it is held to be necessary for a number of other desired and worthwhile things such as creativity, autonomy or self-creation, desert, moral responsibility, suitability of being an object of reactive attitudes, dignity or self-worth, a sense of individuality or uniqueness, life-hopes, freely given love and friendship, and acting of one’s own free will. In “the dialectic of origination,” compatibilists try to “deconstruct” or “demythologize” these things, “to provide plausible explanations of them that do not require underived origination and incompatibilist free will,” while incompatibilists respond by trying to argue that compatibilist versions “fall short of what we *really* want” (80).

To move beyond this impasse, Kane leads us through “the dialectic of self-hood,” where he considers the importance of free will in relation to “our place

or importance in the scheme of things” (92). Far from “delv[ing] rather deeply into the metaphysical depths,” however, Kane instead looks at a most non-esoteric topic in developmental psychology: how a human being arrives at and expands his sense of self. At each stage of the unfolding process of understanding the relation of our selves to the world, we come into contact/conflict with people or things or ideas — including, at one point, the doctrines of determinism — that threaten our experienced status as “independent sources of activity or action in the world,” rather than “mere products of forces coming wholly from the world — forces that are not the products of our own wills” (95). Since we actually desire and believe in this independent selfhood, which Kane says is “a precondition for moral agency in the fullest sense,” we tend to resist anything such as deterministic arguments against free will that could undermine that belief and value (97). Thus, free will is strongly implicated with the idea of independent selfhood, which seems to be a natural, inevitable part of healthy human development; if either is valued, both will be. It seems unclear, however, that anything beyond the compatibilist desires for freedom from constraint, coercion, compulsion, and oppression is required by the desire to be and view oneself as an independent self and actor and willer in the world — which is a relative independence, after all, as Kane himself concedes in his conclusion (213-4). The dialectic of selfhood seems more a continuation and deepening of the dialectic of origination than a new point.

Where Kane really begins to break new ground is in his consideration of “objective worth,” which he suggests is intimately related to ultimate responsibility and free will. When we dig down to the roots of the desire for incompatibilist free will, he says, we find two basic, correlative desires:

- (i) the desire to be independent sources of activity in the world, which is connected . . . from the earliest stages of childhood to the sense we have of our uniqueness and importance as individuals; and (ii) the desire that some of our deeds and accomplishments . . . have objective worth— worth not just from one’s own subjective point of view, but true (i.e., nondeceptive) worth from the point of view of the world. (98)

Kane is saying here that there is more to value than “a person’s subjectively felt happiness” about the results of his acts; and if there is, then incompatibilist freedom with ultimate responsibility *could* have a value beyond that of the compatibilist freedoms: “Such freedoms would be enough, if we did not care about more than what pleases us — namely, if we did not care in addition . . . about our ‘worthiness’ or ‘deservingness’ to be pleased” (97-8). Thus, Kane seems to characterize the debate over a kind of free will “worth wanting,” by depicting the compatibilists as being enmired in a worldly, hedonistic, pleasure-seeking paradigm, in which what they want is most important, in contrast to incompatibilists who want to rise above this shallowness to a more universal, spiritual, “objective worth” paradigm, in which what is *worthy* is most

important. If the proper ethical perspective is from *inside* our own life and awareness, a perspective in which all that matters is supposedly “subjective happiness,” then “objective worth” will seem pointless; but if we “stand back and take an objective view of the universe and our place in it,” Kane says, we will come to believe that our acts have worth “from the point of view of the universe,” which is what he appears to mean by an “objective point of view.”

Kane attempts to persuade the reader on this point by using an “alternative world” scenario, in which a person experiences the same rewards and satisfactions in two different worlds, but one of the experiences is based on reality while the other is a deception. Wouldn’t the former world be “better” to live in, even if no difference was *experienced* between them? Of course it would, but only by virtue of one’s (or someone’s) *projecting* himself into the position of knowing the *truth* about the two worlds; absent such knowledge, all one can go on is the evidence one has. We are not infallible; we may in fact be deceived in one instance or another. But deceit must exist in relation to the truth. If one is deceived and finds out, one can try to adjust. If one never finds out the truth and has a full life of success and happiness in a fool’s paradise, what is the difference — and to whom? Significance is not absolute and cosmic; it is *contextual* and pertains to people living *in this world*. Desiring that there be a cosmic tallyboard toting up one’s true virtues and achievements so that one’s legacy will be accurate, even if humanity’s awareness of it was not, is at best a lapse in self-confidence, a failure to trust one’s own best judgment about whether or not what one has done and made of one’s life and character is good. That some people feel that life is not fully meaningful without this kind of assurance that the universe is “looking out for them” is not a valid argument for incompatibilist free will.

In all fairness, it must be granted that Kane does not attempt in this volume a rigorous justification of his notion of “objective worth” (deferring instead to his *Through the Moral Maze*, Arnouk, NY: M.E. Sharpe, 1994). Even so, it is difficult to see how his proposal is more than just another strategy— and not an altogether necessary one — people might use to help themselves to “resist the idea that the activity we direct back *upon* the world . . . has merely illusory and not real significance or worth *for* the world” (100). In this reviewer’s opinion, however, that idea should *not* be resisted, but instead *embraced*. Kane’s wording reveals rather clearly that he has accepted a traditional false dichotomy between subjective-as-personal and objective-as-impersonal — to wit, that personal values must be subjective and objective values must be impersonal. While personal values often are subjective, they are not always so; more importantly, there is no such thing as “impersonal value,” only the things valued by some *person* (or other living being) for some purpose, life-serving or otherwise.

What this defective dualism thus leaves out is a third view, only recently coming into elucidation: the idea that values can be objective and personal, i.e., factually based and “agent-relative.” (One of the best presentations of this theory can be found in Douglas Den Uyl’s and Douglas Rasmussen’s *Liberty and Nature*, LaSalle, IL: Open Court, 1991. An equivalent perspective can be found in the aforementioned volume by Peikoff, although his relabelling the traditional objective view as “intrinsic” tends to muddy rather than clarify the debate.) Regarding value and significance as not necessarily subjective, but

often instead as agent-relative and factual, allows us to expand the sense of our own personal values beyond the feeling-based, without wandering onto the untenable ground of the point of view of an impersonal universe (into which are often smuggled, without acknowledgment, the *personal values* of someone else and/or the values that someone else who wants to control us wants us to adopt). It is for us and, by extension, those who matter to us (or to whom we matter) — not for “the world” — that our activities have “real significance or worth,” if they have such at all. It is not our “worthiness” to the world that determines whether we deserve happiness, but our *self-worth*. Self-esteem does go beyond the “subjectively felt happiness” (i.e., pride and satisfaction) about the results of our acts, but it is merely a more global, holistic sense of well-being of *ourselves, for ourselves*, not something for which we need the endorsement of an impersonal world. It is Kane’s apparent failure to realize this that leads him to regard the significance of free will as a “metaphysical” issue and part of “the problem of ascent;” and it is this same failure that undercuts his attempt to move beyond the impasse between compatibilists and incompatibilists. In the judgment of this reviewer, Kane has not made his case that there is a noncompatibilist form of free will that is worth wanting.

If there is a case to be made for a form of human freedom that goes beyond the compatibilist freedoms that are subsumable by the traditional event/state, cause-effect model of causality, yet does not lapse completely into the indeterminism that Kane favors for human action, it must pass the test he describes as the Intelligibility Question. This means that such a theory must amount to more than “a perfunctory treatment that consists of putting labels on mysteries” (105), though Kane himself bows to the apparent need to accept such mysteries even for his own view, which includes such notions as “indeterminate efforts” (151) Before fleshing out and defending his own model of “self-forming willings” (124-183), Kane examines a number of traditional and contemporary libertarian strategies to reconcile indeterminism with free will, the requirement being that an agent must be able to willingly do and do otherwise, “all past circumstances and laws of nature remaining the same.” The problem is to explain how a person’s free choice of either of two alternatives satisfies this condition, how one is “able to choose either option rationally, voluntarily, and under [one’s] voluntary control, given the same past and laws of nature.” Proposed solutions to this problem almost always involve some additional factor beyond past circumstances and laws of nature, a strategy that Kane regards as “dangerous,” in that it tempts libertarians to make recourse to “mysterious sources outside of the nature order or to postulate unusual forms of agency or causation whose manner of influencing events is at best obscure.”

In seeking to avoid such approaches involving special forms of agency or causation and to try instead to engage modern science in a “more meaningful dialogue” (115), Kane invokes the Free Agency Principle, which eschews the appeal to entities or causes that are not also needed by nonlibertarian theories: “In the attempt to formulate an incompatibilist or libertarian account of free agency . . . we shall not appeal to categories that are *not also needed by non-libertarian* (compatibilist or determinist) *accounts of free agency* . . .” (116). In the process, Kane weighs and finds wanting a plethora of ideas, including Kant’s noumenal self, Eccles’ “transempirical power center,” the “Will” as a

homunculus or “agent within the agent,” the mind-body dualism of the Cartesian ego, antecedent-cause reasons, and special non-antecedently-caused “acts of will” or “acts of attention.” (Readers of this journal who are familiar with how Leonard Peikoff, Nathaniel Branden, and other proponents of Ayn Rand’s Objectivist philosophy, place the locus of non-determinist free will in the act of focusing one’s awareness or attention, may want to ponder Kane’s assertion that armchair speculation or a priori assertion is not enough to support the claim that such acts are undetermined in principle.)

While he regards such accounts of free agency as risky, however, Kane does not completely reject them out of hand, simply as unnecessary for an explanation of incompatibilist free will; he has his explanation in hand, so these other strategies are at best superfluous. However, I think Kane has missed something vital in at least one of the strategies he mentions: the idea of agent or “nonoccurrent” causation. Kane concedes that his Free Agency Principle only rules out the version of nonoccurrent causation that is designed to account specifically for undetermined, incompatibilistic free actions; but he has little to say about the view held by “only a minority,” viz., the idea that “non-occurrent causation is required to account for action in general, whether determined or undetermined” (122). Edward Pols (see my review of his *Mind Regained* elsewhere in this volume) puts a special twist on this latter idea. His hierarchical model of causality, which puts entities in the causal driver’s seat, agrees that it makes no sense to say that agents engage in immanent or nonoccurrent (“atemporal”) causation in regard to their *actions*, a relation that is fully temporal or occurrent, though not determined by antecedent conditions. He would agree emphatically with Kane’s statement that “[T]he time has come in the history of free will debates when [the] pernicious assumption [that all control must be antecedent determining control] must be subjected to greater scrutiny” (186-7). Pols adds, however, this crucial insight: entities *do* exist in an immanent, atemporal, nonoccurrent causal relation to the parts of their own physical infrastructure that support their temporal actions. This whole-part, downward-upward, pyramidal aspect of causality is, Pols says, what is missing from the microreductionism of modern science’s account of action in general; and invoking the complex arrangement of temporal and atemporal relations in hierarchical causality to help explain human action — whether or not such action is at all indeterministic — would fall well within the strictures imposed by Kane’s Free Agency Principle.

In his further consideration of agent determinism in the brief section dealing with the Existence Question, Kane concedes that “To reject agent-causation is therefore not to deny that there are agents and that they cause things in ordinary senses of the term . . . rather to deny that we need a special relation of nonoccurrent causation to explain all of this” (189). Pols’s way of including nonoccurrent causation in human (and other) action avoids this problem, while showing how it is necessary for fully understanding the hierarchical nature of causality. Kane states that something along the lines of a “recurrent brain network” has to be going on inside an agent’s brain in order to allow an agent’s actions to “outflow” from itself; and similarly to the way I characterized my proposed model of how the brain supports introspection in my review of Pols’s book, Kane says that:

[Such a] “self-network” [need not have] a specific location rather than being a complex network distributed widely throughout the brain. Its unity would lie in the dynamical properties of neural circuits and connections that make possible synchronized and causally interacting oscillations or patterns of firings throughout the entire network, like those described by Crick, Koch, Llinas and others, for conscious awareness and wakefulness. (193)

After the dust settles from Kane’s careful trek through the various controversies about free will and determinism, we are left with a form of indeterminist or libertarian free will that is amazingly like the most tenable versions of soft determinism. Once appeals to “obscure or mysterious forms of agency or causation” are set aside, and a rigorous effort made to formulate an intelligible model of free will and to reconcile it with the scientific view of the world, it becomes clear that the exercise of free will is not “entirely above and beyond the influences of natural causes and conditioning,” and that free will and moral responsibility instead “are matters of degree, and our possession of them can be very much influenced by circumstances” (212-3). Kane thus identifies and rejects for the myth that it is “the idea that we might attain complete autonomy or perfect freedom” (214). Looking to the future, he says that we need to recognize the many ways in which one’s free will and responsibility can be limited by circumstances of birth and upbringing, without being caught in the pitfall of thinking that all of us are inevitably the victims of those circumstances. By rejecting the naive all-or-nothing views of free will, we can excuse or mitigate the guilt of those with diminished moral responsibility due, for instance, to severe childhood abuse, while (with few exceptions for extreme cases) not letting such people completely off the hook. As for the future of the continuing intellectual debate, it is safe to predict that the framework provided in Kane’s book will be adopted by many as the best guide for delving into the intricacies and nuances he has outlined.

Roger E. Bissell

## Book Reviews:

### *Bouwsma's Notes on Wittgenstein's Philosophy, 1965-1975*

J. L. Craft and Ronald Hustwit, editors. Lewiston, NY:  
The Edwin Mellen Press, 1996.

Wittgenstein says in *Philosophical Investigations* §133 that

The real discovery is the one that makes me capable of stopping doing philosophy when I want to. —The one that gives philosophy peace, so that it is no longer tormented by questions which bring *itself* in question.— Instead, we now demonstrate a method, by examples; and the series of examples can be broken off.—Problems are solved (difficulties eliminated), not a *single* problem.<sup>1</sup>

As I understand this passage, Wittgenstein does not claim that in the *Investigations*, he puts an end to philosophy. Rather, he tries to eliminate certain philosophical difficulties. He does this not to end philosophy but to demonstrate a method that we can use in trying to eliminate those and other philosophical difficulties. So, according to one reading of *PI* §133, there are philosophical difficulties that Wittgenstein does not address in the *Investigations*; he breaks off “the series of examples” that he uses to demonstrate his method. He does this at least in part because he wants us to learn to use his method, and good teachers never solve all of their students’ problems; good teachers never “spare other people the trouble of thinking.”<sup>2</sup> We might say, then, that Wittgenstein leaves us a few practice exercises that we can use in completing our homework assignment, that is, in learning to use his method.

From the 1930’s, when he first encountered what we now know as Wittgenstein’s *Blue Book*, until his death in 1978, O. K. Bouwsma was a student of Wittgenstein’s; he was during that time doing the homework that Wittgenstein assigned. Bouwsma constantly used Wittgenstein’s method to try to solve the philosophical difficulties that faced him. And since using Wittgenstein’s method helps us not only to solve philosophical difficulties but also to understand more clearly exactly what that method is, Bouwsma’s work was a constant effort both to learn to use and to gain a deeper understanding of Wittgenstein’s method.

Bouwsma’s *Notes on Wittgenstein's Philosophy, 1965-1975* represents the efforts that he made late in his life to use and to understand Wittgenstein’s method. We might put it this way: Bouwsma’s *Notes* is made up of some of his last filled-in homework notebooks. In the *Notes*, Bouwsma practices using Wittgenstein’s method when addressing a variety of philosophical problems that arise from the comments of Bouwsma’s students and from passages in the writings of philosophers. The subjects of these comments and passages range

from time and the self to the nude in art and the categorical imperative. His treatments of these subjects are thorough and detailed, but in many places they come complete with what Bouwsma himself considered failures (p. 81), defects (p. 251) or diversions (p. 127). Yet in spite of Bouwsma's failures, defects and diversions, or probably in virtue of them, reading his *Notes* is like watching a good student do his homework, and watching him do his homework is tutorial—it helps us to do our own homework.

To do our homework, to learn to use Wittgenstein's method, we must realize that we use our words in countless ways (*PI* § 23), and we must be able to recognize at least some of the different ways in which we use our words. As Bouwsma says, we must "cultivate the art of discerning differences among our uses of language and the skill of describing the varieties of such uses" (p. 369). An important part of our homework is to cultivate that art, to train our ears to recognize uses, as well as abuses and misuses, of language.

In the *Notes*, Bouwsma tries to train our ears and, thus, helps us to be able to recognize abuses and misuses of language. He sometimes play[s] on your ear with wrong descriptions and with wrong words and discords in the language. This is for practice. Such discipline as is required for this, a good ear, cannot be taught as 1, 2, 3. It takes time and incessant practice. (p. 369)

Bouwsma plays on our ears with discords in the language, not to mislead us, but to train us to recognize such discords. Bouwsma has us consider, for example, the expressions in the following list (see p. 292):

- "The meaning of"
- "The spelling of"
- "The pronunciation of"
- "The history of"
- "The etymology of"
- "The first syllable of"

For the most part, the expressions on this list are expressions that we understand. Bouwsma claims, however, that the first of these expressions—"The meaning of"—is out of tune with the others. He says, "Why should '*What is the meaning of a word,*' strike one dumb? Whereas, '*What is the spelling of a word,*' would not?" (p. 293). When we compare "The meaning of" with the other expressions on Bouwsma's list, we can see that the way we use "The meaning of" is different from the way we use the other expressions. In fact, according to Bouwsma, comparing bits of language is essential to recognizing discords in the language. He wants to show us that if we are to recognize that expressions like "The meaning of" are discordant, we must examine them in the light of a variety of similar expressions. For just as we cannot hear that C-sharp is out of tune with other notes if we listen only to C-sharp, so too we cannot see that "The meaning of" is out of tune with other expressions if we examine only "The meaning of."

Why, though, is it important to be able to recognize these discords? According to Bouwsma (and to Wittgenstein), when we fail to recognize that a particular expression is out of tune with others—that is, when we mistakenly

believe that certain expressions are used in the same way as other, perhaps similar expressions—philosophical difficulties can arise. The mistaken beliefs that cause these difficulties are what Wittgenstein might call mistaken analogies between different forms of expression (*PI* §90). Bouwsma claims, for example, that difficulties arise when we believe that “red sensation” or “red image” is analogous to “red tomato,” “red crayon,” or “red apple” (see pp. 79-81). He says,

Notice if I ask, “A red tomato,” you can show me the tomato. “Indeed red.” But if I ask, “The image of a red tomato?” You have nothing to show for it; you are helpless with your image, if you have an image. And I, of course, I may feel even more helpless since I cannot even have a look for myself, as you can. (p. 80)

Bouwsma points out that if we give in to the temptation to believe that “red image” is analogous to “red tomato,” we will probably accept “red image”—a bit of disguised nonsense—as sense. And if we accept as sense bits of nonsense, we will soon enough feel philosophically helpless.

Bouwsma tries to show us how, by using Wittgenstein’s method, we can avoid such philosophical trouble. He tries to show us that we will fail to understand expressions like “red image” if we insist that they are analogous to expressions like “red tomato.” This mistaken analogy disguises “red image” as an expression that makes sense and, thus, as an expression that we should try, but that we are doomed to fail, to understand. Bouwsma uses Wittgenstein’s method to remove the masks of sense from that bit of disguised nonsense. And Bouwsma wants to help us to be able to remove these masks for ourselves, to recognize for ourselves that certain bits of disguised nonsense that we accept as sense are, in fact, not sense at all.

To that effect, Bouwsma sometimes “translate[s] or elaborate[s] [on a bit of language] in order to bring the sense or the non-sense to light—usually the non-sense” (p. 138). For example, Bouwsma says,

I get “the sensation of red,” an object of perception (I see it), the sensation, and I call it red. The image of red serves in the same way. It will then follow that when I call “it” red you cannot see “it” and when you call another “it” red I cannot see “it,” so you cannot understand me and I cannot understand you. (p. 78)

This elaboration suggests that “red image” and “red sensation” are non-sense; if we say that our images or sensations are red, we say things that cannot be understood. Bouwsma goes on to point out that unless we see expressions like “red image” and “red sensation” as nonsense, we are tempted to try to understand those expressions. Yet we encounter the philosophical difficulties that come with trying to understand those expressions only if we try to understand them. So removing the temptation to try to understand them—rec-

ognizing “red image” and “red sensation” for the nonsense that they are—prevents the philosophical difficulties that result from trying to understand those expressions.

Few of Bouwsma’s elaborations, however, are as brief as the above quotation. Usually, when he translates or elaborates on a bit of language, he takes no shortcuts, if there are any that he may take. Instead, he takes the long way home. He worries bits of language “in the same way that a dog worries an old shoe” (p. 122), deliberately gnawing, searching for the chewiest parts, tugging at it from every angle, rolling it around in his paws for most of the day, until the leathery cows come home. He does this in order to see—or to hear—more clearly which uses of an expression are in tune and which are out of tune. Bouwsma’s worryings let us see, not only when we understand certain uses of a word, but also when we fail to understand uses of that word: His worrying with “red” helps us to see not only that many sentences in which “red” is used make sense but also that sentences like “My image of a tomato is red” make nonsense.

At times, though, Bouwsma seems to worry a bit of language too much, for he sometimes gets a bit lost or, perhaps better, diverted. We sometimes find him, and he finds himself, distracted by an ultimately less helpful worrying. At times he is like a geologist who, during the course of collecting a variety of stones, stops at the pond to skip a few. (“Language is our delight,” Bouwsma says, “as well as our workhorse” (p. 82).) For example, Bouwsma worries “red” in the following way:

Blood, her lips, my sweater, a beet, sunsets, the big chair. It is in the surroundings of such things that we learn the use of the word “red.” “Today you will wear your red hair ribbon and your red socks.” “I have a sister who has red hair.” “There’s the red flag and there’s the red, white, and blue.” Roses are red. Peaches have red cheeks. Fire departments trucks are red. Coals in the grate are red. Tomatoes are red, and so are some apples. . . . (p. 79)<sup>3</sup>

Perhaps diversions such as these are grounds for criticism of Bouwsma’s *Notes*.

On the contrary, these diversions are welcome, both because they are delightful and because they can help us to do philosophy in the way that Wittgenstein would have us do philosophy (or, perhaps better, because they can help us to avoid doing philosophy in ways that Wittgenstein would have us avoid). Wittgenstein claims that when we are in the mi(d)st of a philosophical problem, a main cause of our difficulty is limiting ourselves to only one kind of example of the use of a particular bit of language. We have seen that, according to Bouwsma, certain philosophical difficulties arise when we examine “The meaning of” or “red sensation” in isolation from other, similar expressions. Limiting ourselves in either of those ways—which would be, in effect, to allow ourselves only a sort of tunnel vision—is dangerous because when we are so limited, we tend to answer our questions, to solve our philosophical problem,

on the basis of that example alone. But we ought not overlook other examples, for they might call into question the answers that we gave when we focused on only one example. Other examples are helpful because they might allow us to see that the answers that we gave when we focused on only one example are incomplete or, what is worse, confused.

Typically, Bouwsma's worryings, even those that are diversions, forcefully and gracefully remind us of the variety of examples of use of language, for we should count among that variety even those examples that Bouwsma's diversions produce. And since Bouwsma supplies himself with a variety of examples, he need not nourish himself with only one example; he can avoid "a one-sided diet" (*PI* §593). If we learn from Bouwsma to worry bits of language as he does, we too will be able to provide ourselves with a variety of examples and to avoid the troubles that can arise from focusing on only one example. Bouwsma's worryings remind us to remind ourselves, and they show us how to remind ourselves, of the variety of kinds of use of language.

Furthermore, once we remind ourselves of the variety of kinds of use of language, we will be in a better position to understand and to eliminate certain philosophical difficulties. Bouwsma says,

If in any philosophical discussion you have trouble with a word [and] you cannot understand it in that context then to gain perspective[,] you may set it in a context in which you do understand it. In that way, you may discover [w]hat is wrong when you did not understand it. (p. 175)

If we worry bits of language as Bouwsma does, even if our worrying sometimes includes diversions, our chances are increased of finding a use of some bit of language that makes sense and of revealing uses of that bit of language that make nonsense; the more we worry a bit of language—that is, the more we play "with the grammar or with the sense" (p. 228) of a bit of language—the better chance we have to discover which uses of that bit of language we understand and which we do not. In turn, worrying bits of language as Bouwsma does increases our chances of discovering what is wrong when we do not understand some use of a word and, thus, of discovering what our philosophical troubles are.

When Bouwsma, in the *Notes*, tries to see how philosophical difficulties arise and how, by using Wittgenstein's method, to eliminate those difficulties, he is doing the homework that Wittgenstein assigns us. By watching Bouwsma do his homework, we are, I think, in a better position, we are better equipped, to do our own homework. Bouwsma's *Notes*, if we use them as a supplement to the writings of Wittgenstein and as an example, can help us learn to do philosophy in the way that Wittgenstein would have us do it.

But J. L. Craft claims that Bouwsma's *Notes* are more than a filled-in homework notebook: Craft claims in his introduction to the *Notes* that they "are a kind of commentary on certain themes in Wittgenstein, significant themes in *The Blue Book* and *P. I.*" (p. xi). But on what themes are Bouwsma's *Notes* sup-

posed to be a commentary? As I see it, Craft might answer that question in any of three ways. First, he might say that Bouwsma's *Notes* are a commentary on the use of Wittgenstein's method. Now, to use Wittgenstein's method, we need, among other things, to get clear on certain things that Wittgenstein says in the *Investigations* and in *The Blue Book*. To that effect, Craft says that "Bouwsma wanted to bring out [certain] key themes or 'discoveries' of Wittgenstein's" (p. xii). As I have already suggested, if Bouwsma's *Notes* are meant to be a commentary on the use of Wittgenstein's method, where part of being able to use that method is trying to understand certain themes or passages in Wittgenstein's writings, then Bouwsma's *Notes* are a valuable and illuminating commentary.

However, Craft might say that Bouwsma's *Notes* is a commentary on certain problems that Wittgenstein himself addresses in his writings. Bouwsma spends time worrying not only over problems that Wittgenstein does *not* address but also over some problems that Wittgenstein does address in the *Investigations* and in *The Blue Book*.<sup>4</sup> Craft might say, then, that when Bouwsma worries over some problem that Wittgenstein worries over, Bouwsma is giving a sort of commentary on Wittgenstein's treatment of that problem. However, it is much more useful to think of Bouwsma's worrying over those problems not as a commentary on Wittgenstein's treatment of those problems but as Bouwsma's own effort to eliminate some of the same difficulties that Wittgenstein himself tries to eliminate. With Wittgenstein's guidance, Bouwsma struggles with some of Wittgenstein's problems, yet we should not, I think, represent Bouwsma's struggles as a commentary on Wittgenstein's struggles.

Last, Craft might say that Bouwsma's *Notes* is at least in part a commentary on specific sections of the *Investigations* and on specific passages in *The Blue Book*. But this, too, is the wrong way to think of Bouwsma's *Notes*; it is best not to see the *Notes* as a commentary in the sense of explanations or interpretations of Wittgenstein's remarks. Bouwsma himself admits as much. For example, in the course of his discussions of §43 of the *Investigations*, Bouwsma says, "What I wrote must not be taken as an account of how W [Wittgenstein] understood 43, how he intended it" (p. 202). Also, Bouwsma at times admits that he has failed fully to understand some passage or other of Wittgenstein's. He says, at the end of his notes of 21 March 1968, in which he ponders PI §273 and related sections, that "I am still troubled by 'the sensation of red'" (p. 81). At least on these occasions and, I think, on many more, Bouwsma did not intend his notes to be a commentary on specific remarks in the *Investigations* or on specific passages in *The Blue Book*. We should not, then, look upon Bouwsma's *Notes* only as a commentary on specific themes or passages in Wittgenstein's writings.

Since "Bouwsma did not compose [his] notes for publication and never intended their circulation among a wider audience than his students" (p. xi), we might think that the publication of Bouwsma's notes is a disservice to Bouwsma and perhaps even to the readers of the *Notes*. Yet in the end, if we see the *Notes* for what it is, as what I have called a filled-in homework notebook rather than as a commentary, then Bouwsma's *Notes* can be inviting, instructive, and, at times, downright fun. In the *Notes*, Bouwsma entices us to ponder certain philosophical problems, and he helps us see how we can use

Wittgenstein's method when we ponder those problems. But Bouwsma does more than that in the Notes, for he also reminds us of the joys of skipping stones, of gnawing on an old shoe, of delighting in language.

### Tim Black

1. Ludwig Wittgenstein, *Philosophical Investigations* (New York: Macmillan Publishing Co., Inc., 1958).
2. Wittgenstein, *Philosophical Investigations*, p. vi.
3. Compare Bouwsma's worrying with the word "word" (pp. 113-5). For that worrying, he was criticized by a student: Bouwsma says, "In connection with examples of sentences containing the word 'word,' Lee Gordon remarked about the surfeit of examples. Four or five would be enough. He is no doubt right about that. Enough is enough" (p. 127).
4. Bouwsma worries, for example, about the notion of a private experience (see, e.g., pp. 75-79), about which Wittgenstein worries, and about the nude in art (see pp. 417-420), about which Wittgenstein did not worry (or, at least, not in *Philosophical Investigations*).

## Book Reviews:

### *Animals and their Moral Standing* by Stephen R.L. Clark. New York: Routledge, 1997.

This collection of self-selected essays, all previously published (with some revision), presents work of Clark's from 1978 through 1994. The essays are engaging, often poetic, but philosophically demanding. They illustrate Clark's unique approach to the animal welfare debate. A metaphysical realist about moral and mental properties, Clark fights the current trend toward skepticism and relativism with a Moorean common sense while urging the reader to recognize the valuable epistemological role that human sentiments (moral and otherwise) play in justifying the judgments that we make about our moral community. Clark's essays invite the reader to engage in the moral exercise of recognizing the role that creatures play in our moral lives and the trustworthiness of the sentiments that lead us to invite them into our household. These are the sentiments that form the foundation of the cosmic democracy that Clark envisions. Clark brings an honest humility to his writing as he tackles a difficult subject, and, writing from a Christian perspective, creates for himself a difficult (because diverse) audience. Moreover, his ability to bring together contemporary sociobiology, ancient and modern philosophical perspectives, and theological perspectives is quite remarkable. The result is a collection of essays that will be of interest to the veteran as well as the newcomer to the animal welfare debate.

The most recent of the essays "Modern Errors, Ancient Virtues" provides a nice summary of the ideas that unite many of the essays in this collection. Clark reminds the reader that those engaged in morally questionable practices using non human animals operating on a set of false assumptions deeply ingrained in contemporary (human) culture: Egocentrism, the view that the world is essentially our construct and we are the center of that world; Humanism, the view that the only species deserving of moral respect is the human species, and this, because its members are the unique embodiment of reason; Utilitarianism, the view that while non human sentient beings might deserve some moral respect, the good that comes from ignoring them is justified in the name of the overall good; Objectivism, the view that we regulate our feelings and behavior not by conventional and changing concepts but by the natural or real division of things.

The other essays in the collection offer more elaborate, suggestive attacks on some of these assumptions.

"How to Calculate the Greater Good" and "Ethical Problems in Animal Welfare" may be the least interesting for veterans of the field, but most helpful for newcomers. Here Clark presents and critically examines the key concepts and normative theories that have defined the animal welfare debate. His critique of Utilitarianism, based on its epistemological inadequacies, sets the stage for his appeal to a "cosmic democracy" established and reinforced by our emotional, moral responsiveness as human animals. "Cosmic democracy" the

idea that the moral sensibility that binds us to our children and pets must be extended to the whole earth, is further elaborated in "Utility, Rights and Domestic Virtues" and "Animals, Ecosystems and the Liberal Ethics."

"The Rights of Wild Things" and "Hume, Animals and the Objectivity of Morals" also provide a wonderful introduction to some of the key issues in the debate and Clark's realism about morality concerning the animal kingdom. In the former Clark examines D.H. Ritchie's reductio that granting natural rights to animals will force us to the ridiculous conclusion that we must protect animal prey against predators who wrongly violate the victim's rights. Clark responds by essentially embracing Ritchie's conclusion and showing the reader why common sense and moral sentiment support this view. Emotional responsiveness to the life and pain of human animals warrants our protection of both human and non human animals, especially against predators who prey not out of necessity but enjoyment. The strength of our commitment is a function of their membership in our household - a membership warranted by the advantages we gain from them. However, no one would require someone to protect a right if it violated other duties, or made things worse by protecting it, or interfered with another whose predatorial actions were based on need not greed. So while rights may obligate us to protect, it doesn't obligate us to protect unconditionally.

The most suggestive and bold idea that emerges in these two essays is that a predator's actions are susceptible to moral evaluation. This thesis emerges partly in response to Hume's challenge that a moral realist cannot consistently treat non animal incest as the same act as human incest, while making different moral attributions. Clark, in keeping with his view that we can attribute a mental life to animals, suggests that the animal's action is condemnable - even if the animal does not recognize it as such.

Another group of essays illustrate Clark's attack on the implausible assumptions that he believes have a grip on the philosophical, literary, and scientific establishment: that we need to distinguish between the factual or natural world that science investigates, and the value-laden, prescriptive judgments that we make on the basis of our emotional response to it. Clark's attack involves two related theses, one metaphysical, one epistemological. He defends the epistemological thesis, namely that we do not infer the presence of these mental capacities and moral ties, but rather directly see them in virtue of our emotional responsiveness as animals and loving attention as scientists most clearly, in "Awareness and Self-awareness," "Humans, Animals and Animal Behavior," and "The Description and Evaluation of Animal Emotion." In the essay "Awareness and Self-awareness" Clark argues that linguistic self-ascription is not the only sign of self-awareness. As animals we are equipped to recognize self-awareness; we do this not through inference but by identifying the same non-linguistic behavior that we use to identify self-awareness among humans. For example, we observe how one locates oneself in space, re-identifies individuals as individuals instead of mere occupants of a role, and we observe the degree of commitment and admitted responsibility to others. In addition to awareness, animals have perceptions, projects, and interests all of which form the basis for their membership in the moral community. In "Humans, Animals and Animal Behavior" Clark examines the traditional con-

nection between animal welfare and asceticism where animals gain protection on the basis of disassociating reason from passion or animal concerns. Clark's argument is that any morality that systematically denigrates and denies its own roots in ethical responsiveness is doomed. In "The Description and Evaluation of Animal Emotion" Clark further defends the epistemological role of human responsiveness by examining Spinoza's challenge that emotion in animals is essentially different from humans and Howard Liddell's use of the descriptive/evaluative distinction to emphasize the limits of emotion attribution to animals. Clark's response is to address the implications of ignoring our sentiments: either we detect a part of the nature of the animal through our response to their "manifestation of emotions" or these responses are defective because their mental lives are completely alien to us. But their mental lives are not alien to us. We train them, we communicate with them sufficiently enough to feel comfortable with our interactions with them. This argument, while compatible with the fact that we could be mistaken, is reason enough to grant that our responsiveness provides knowledge about some aspect of their nature.

The epistemological role of our responsiveness appears again in the essays where Clark defends a metaphysical realism about folk psychology and morality. In "The Reality of Shared Emotions" and "The Consciousness of Animals" Clark addresses two standard views, that either reject the truth of our prescriptive judgments (material eliminativists) or reduce the truth of such judgments to a matter of what we are willing to countenance at the time (anti-realist or post-modernists). Both, as Clark rightly points out, deny that our moral and psychological concepts have a hold in reality. Clark attacks both views on three fronts. 1. Both views ignore the history of our concepts: our concepts are grounded in reality in virtue of our interacting with the world, as human beings. Consequently, while some may be less appropriate than others, conventional "value laden" concepts are as much a part of that reality as the "natural" or "scientific." 2. The skeptic's claim that even if we interact with the world, we may not have an appropriate understanding of the concepts as they apply to non humans is no reason to think that there is no way to apply them, or that we can only apply them metaphorically. We may very well never know what it is like to be a wasp, for example, but functional similarity is enough to bring one into the moral universe. 3. Not only is impersonal science impossible at a global level (scientists for some time have admitted the impossibility of eliminating all reference to an inner life in order to describe and explicate animal behavior), but it may be scientifically inappropriate if the project of discovery is taken seriously. For our emotions are not merely mechanisms of "projections" but can discover something that impartial reason may not. As Clark repeatedly reminds the reader: our success at training animals by acknowledging the importance of emotional reward and punishment seems evidence enough.

Clark's essays are intellectually demanding and rich with philosophical creativity, insight, and argument. Of course there is much to question and challenge as a result of this richness. For instance, Clark's appeal to the notions of "natural law" and of an animal's "natural life" in order to defend why we must not only refrain from cruelty, but killing as well, raises serious questions. For instance, can autonomy or desire for sacrifice or concern for suffering override

that right to adhere to its nature? Can its nature be determined by our conventions (e.g. breeding laboratory mice for research)? Are we allowed to change its nature for the benefit of genetic research? I can imagine two other obvious challenges that might emerge from opponents and sympathizers. I think some zoophiles may well be frustrated with these essays feeling that he has not done enough. While he most certainly addresses the skeptics and those who are not willing to recognize, as he puts it, the “easy duty” to refrain from violating rights, he doesn’t do much by way of guiding us with the more difficult duties, settling those conflicts we confront once we recognize a cosmic democracy, telling us what sympathies to trust, what moral discriminations or “exclusions” we must make. But I suspect that Clark will simply argue that this problem is no different from the problems faced by the speciesist when dealing with conflicting duties and sympathies involving humans of differing capacities and interests. As for the difficult duties, I suspect he would offer an Aristotelian view: who one helps and to what extent and when, will depend on the nature of the individual and his or her capacities and position.

One objection I anticipate from philosophers, regardless of their sympathies, is that Clark fails to address more sophisticated forms of Utilitarianism and Emotivism. While this is true, the kind of strategy he employs may cut against both traditional and more sophisticated versions of these views. Indeed, what I enjoyed most in Clark’s responses to his opponents is what some readers will find annoying, and most certainly question-begging: his common sense responses to skepticism and relativism. While this collection of essays will serve the student interested in the particular topic of animal welfare, it is also a brilliant illustration of common sense philosophy in the tradition of Aristotle, Thomas Reid, G.E. Moore, and, most recently, John McDowell. Anyone who questions the contribution of traditional philosophy to modern ethics or current science would do well to spend time with this book.

Jody L. Graham