

CAMPBELL'S BLIND VARIATION IN THE EVOLUTION OF AN IDEOLOGY AND POPPER'S WORLD 3

Ray Scott Percival

Introduction

My general problem is the perpetuation and reproduction of systems of ideas. What can be done to systems of ideas, how can one tinker with them, to enhance or reduce their criticizability and copyability? Can a system of ideas be turned into what Richard Dawkins has called a "Mind-Virus", something that is retained and copied in the face of all possible counter-evidence? Dawkins (1976, p. 198) and many others think so.

William Warren Bartley's general position on the openness to criticism of ideologies is that

ideologies are retained regardless of the facts; they are not abandoned when they clash with the facts; rather they die out or are eliminated, if at all, together with their carriers... (1962, p. vii)

The claim is that there are networks of theories making certain claims about the world whose proponents continue to maintain and propagate them whatever facts are presented against them. This view is reminiscent of Planck's view of science. Planck held that new theories in science become accepted only because the proponents of the old theories die off, leaving it to the young generation of scientists to adopt and develop the new theories.

It is helpful to see the problem in terms of the determined propagandist, prepared to use deception and force, and whose two goals are:

- (1) to guarantee the propagation of his doctrine and
- (2) to guarantee it from criticism.

I argue that neither goal can be fulfilled, and that they must be traded for one another: maximizing the copyability of a doctrine means making it more open to criticism; maximizing its closedness to argument sacrifices some of its copyability.

I think that the belief in incurable "mind-viruses" is popular partly because of the dominance of materialism. From this perspective copying ideas is rather like photocopying - a physical process that is rather transparent and easy to control. But Campbell (1979) taught us that when faced with a problem in communication we should look at the distinctive structural and vehicular characteristics of the message. Every message, whether a poem, painting, mosaic or a scientific theory, has to be embodied in some vehicle for transmission, and this vehicle has characteristics that will frustrate perfect transmission in distinctive ways. The very community in which this message has significance, is itself, Campbell emphasises, a vehicle or carrier of the message, and this community will also aggravate the problem of accurate transmission. I will argue that the logical characteristics of "mind-viruses" make them impossible to fully control and ensure their reproduction. A more accurate analogy than photo-copying would be the attempt to maintain a lie. In order to maintain one lie in the face of awkward questions a hundred have to be created. But this extra theoretical baggage becomes less easy to maintain, and even more difficult to spread faithfully through a population.

The general thrust of my thesis is that humans are rational and open to argument. This is true of supposedly closed systems of ideas in science, religion and politics and their adherents, and I try to show how they are, despite appearances, open to argument and counter-evidence. Truth, acting through the falsification of theories, is what Campbell (1993) would call a "coselector" of ideologies or "mind-viruses". I will elaborate on and defend the relevance and force of Popper's notion of logical refutation as a coselector of ideas, both in- and outside science.

Essential to this analysis of openness to argument is Campbell's "variation-with-selective-retention" evolutionary schema. This schema has been applied in many fields very productively. But an equally essential element in my analysis is Popper's notion of World 3, the world of abstract products of the human mind but not reducible to it (Popper &

Eccles 1977). An example of a World 3 object would be the natural numbers. They may have been invented originally simply for the purpose of counting. However, once created, they were found to have characteristics autonomous from and no part of this original purpose, properties such as the prime number sequence. Campbell saw a selective role for abstract logic as an internal selector:

If there are such logical truths about inexorable limitations on such inference rules, such logical truths are a part of the selective system editing the mutations which introduce variations into inference and decision-rule anatomy. Thus an abstract analysis of the logically possible can describe part of the environment being biologically adapted to. (1989?)

In this truly marvelous paper, Campbell tries to physicalize the three evolutionary stages, and this is where my reservations lie. The role of abstract relations and objects cannot be fully physicalized. I consider an ideology - whatever else it might be - to be an abstract product of the human mind, a theoretical system. Like other World 3 objects, an ideology has abstract properties and relationships that go beyond the psychology of the ideologist. But these properties are nevertheless real and can have real effects in both the psychological world and the physical world. It can be shown that theories and hence ideologies, have an unfathomable content. Their logical implications and ramifications are not fully known to their authors. Once created, ideologies have a life of their own, and can surprise, shock and frustrate the author of the system.

In other words I am assuming that such things as logical relationships (such as validity and invalidity, inconsistency etc.), make a difference in the evolution of an ideology. Thus when analysing the evolution of ideas, World 3 objects must be incorporated into Campbell's "Variation-with-selective-retention" schema. Although at odds with Campbell's quest to completely physicalize his schema, the unfathomableness of ideologies or mind-viruses reinforces Campbell's point about the blindness of the evolution of systems of ideas. Obviously, a determined propagandist will try to impose his design for the spread of the "mind-virus", but an irreducible, profound blindness is encountered at all three stages in the evolution of a "mind virus":

1. Blind Variation:

To spread his message, the propagandist creates an embodied version of his idea whose objective content exceeds his personal knowledge. Further blind variation from other sources maintains an ocean of freshly created, unfathomable rivals.

2. Blind Selection:

The idea enters an ocean of rival ideas. The logical interaction with these rivals cannot be fully known. But logical criticism from other ideas is one of the coselective filters of ideas. The propagandist, therefore, cannot prepare comprehensive defences in advance. Moreover, how each defence will effect the next round of selection also cannot fully be known.

3. Blind Reproduction:

To reproduce an idea, other ideas must be recruited to help. Exactly which other ideas will be required to reproduce the idea cannot be fully known in advance.

This approach contrasts with an approach which assumes that ideologies are simply expressions of the psychology and sociology of a community, as if they were domains isolated from the effects of logic. My approach allows us to analyse the interaction of the very different domains of psychology, sociology (Popper's World 2) and methodology and logic (Popper's World 3). The analysis shows that the achievement of what Campbell calls a "Self-Reproducing Social System" is almost impossible, fundamentally because the profoundly unpredictable products of the three stage evolution of World 3 entities are constantly disturbing any attempt to impose a strict copying of the original message. Any slight deviations can accumulate, and even without accumulation, small variations in a theoretical system can radically alter its ramifications.

A ubiquitous assumption is that an ideology can set up logical barriers to criticism, giving it an evolutionary advantage in the competition of ideas. The possibility of setting up such logical barriers to criticism can be explored in connection with the so-called immunizing stratagem. I shall also explore the implications for the propagandist of Duhem's and Lakatos's argument that a theoretical system can only be tested and criticized as a whole because any part of it may be made safe from criticism by suitable adjustment to other assumptions of the system.

Many immunizing stratagems involve abandoning the ideology for whose protection they have been introduced, an unplanned, often unforeseeable, process that consists of numerous successive slight

modifications extending sometimes over hundreds of years. Other immunizing stratagems seriously lower the survival value of the ideology through the acquisition, sometimes over a long period, of a burdensome and confusing "protective belt" of hypotheses, each of which acted at least in the short-run, to deflect criticism away from a privileged sector of assumptions.

Duhem's and Lakatos's arguments cannot be used to show that a privileged part of a system may be guaranteed from criticism because of the unfathomable content of theories and hence the unforeseeable evolution of their defence. I will expand in detail how this leads to the emergence of factions that disagree on how best to defend an ideology and even on what to count as the privileged part of an ideology and what to count as the dispensable part.

Karl Popper originally used the term "conventionalist stratagem", but then adopted the term "immunizing stratagem" from Hans Albert to describe an aspect of the unscientific methodology of certain ideologies claiming to be scientific, Marxism and Freudianism. Apparently Arther Pap anticipates this usage.¹ Popper argued that Marxism, which originally was an empirically testable theory, had been recast in the form of empirically irrefutable metaphysics. This manoeuvre, Popper argued, saved Marxism from refutation and immunized it against further attacks (1976, p. 43). Freudianism was, Popper says, irrefutable from the beginning. The basic theory of Freudianism does not need any immunization to make it irrefutable. Nevertheless, it does incorporate immunizing stratagems. Popper contrasted these two theories with the theories of Newton and of Einstein which were full of testable (i.e. falsifiable) content. Thus the term "immunizing stratagem" arose in connection with Popper's attempt to solve the problem of distinguishing scientific from pseudo-scientific theories - the so-called demarcation problem. Popper's solution was the methodological rule to allow into science only empirically falsifiable hypotheses, and subject these to severe criticism. In addition, theory development was to proceed from less to more testable, i.e., more informative theories. If a theory is refuted and an alternative sought, it had to be more testable, not less, and the more testable the better. For to reduce testability is to reduce knowledge, but in science we desire the growth of knowledge. An immunizing stratagem is a development in theory that reduces testability.

Not all evasive moves are on the wrong side of the demarcation

criterion. Some auxiliary hypotheses introduced to deflect a refutation from a valuable assumption have added greatly to our knowledge. One such auxiliary hypothesis was the prediction by Adams and Leverier of the existence of the planet Neptune.

My point, contrary to Popper, is that "immunizing stratagems" are auxiliary hypotheses that are on the wrong side of the demarcation criterion and precisely those that while saving the original theory from refutation effectively abandon it, replacing it with another theory.

1. Popper's Own Examples of Immunizing Stratagems

Popper says that immunizing stratagems save theories from refutation. However, Popper's examples are not examples of saved theories but examples of repudiated theories: to immunize a theory in these cases is to abandon it. The two main effects of these immunising stratagems are (1) Saving the theorist from embarrassment at the price of abandoning the original theory; (2) clouding the issue and reducing information content. I think that Popper was dimly aware that immunizing stratagems do not strictly save theories (in some cases he puts the word "saved" in scare quotation marks), but he did not see the full implications of this, especially for the survival of an ideology.

To illustrate points (1) and (2) I have chosen the simplest of Popper's examples. Popper (1934) asks us to consider the case of a man who makes the bold claim that all swans are white, but on being presented with a black swan promptly denies that it is a swan. After all, he says, whiteness is part of the definition of the word "swan". Popper says that the theory has been saved from refutation. However, what had been an empirical theory about the world was turned into part of a definition. This is more accurately expressed this way: the original theory, supposedly protected by the immunizing stratagem, has been replaced by an implication of a vacuous definition.

This logical point is worth expanding. The original theory was empirical in Popper's sense: it was capable of clashing with reality. The statements "All swans are white" and "There is a black swan" cannot both be true. A definition or implications derived exclusively from a definition, however, cannot clash with reality for they say nothing about the world. Thus there could not have been a more drastic repudiation of

the original theory: only the words are the same. But the repudiation is implicit and unacknowledged, thus saving face despite abandoning the original claim. Once this is accepted we can derive some interesting implications about the evolution of an ideology under criticism.

Of course, in a real life people do not simply make such bold assertions out of the blue. Rather, they are made with a certain intention, background assumptions and more or less clearly formulated problems. It is this context of assumptions and problems that both guides us in identifying an immunizing stratagem and in refuting the original assertion. For example, the sentence "All swans are white" might be derived from a biological theory of colouring in birds. Knowing this allows us to exclude a whole range of immunizing stratagems that contradict this biological theory or seem to make irrelevant the intention of maintaining the biological theory as a solution to the problem of colouring in birds.

Provisionally, we may define an immunizing stratagem as an evasion of falsification by the reinterpretation of a theory or its ostensible problem, the modification of its assumptions, or by the reinterpretation or simple denial of the counter-evidence, so that the modified theory is then consistent with the evidence.

I see the use of immunizing stratagems not as a sign of an ideology in Bartley's sense, as a complete disregard of truth, but rather of a confused and incompetent attempt to take account of criticism. Those resorting to immunizing stratagems are rather like the American Officers in Vietnam who said that they had to destroy a village in order to save it. Thus I also disagree with Anthony Flew. Flew characterizes evasions of falsification as involving "surreptitious" and "arbitrary" manoeuvres (1975, p. 48). They also show "that your concern is with what you would like, rather than with how in truth things are" (p. 54). My argument is that the changes may not be designed, but may be the unintentional consequence of an attempt to deal with criticism and retain the theory. To the extent that the manoeuvres abandon the original doctrine in response to the specific falsification involved they cannot be wholly arbitrary. This comment reinforces my point that falsification can act as a Darwinian-like filtering device on ideologies even if evasive (intentional or unintentional) moves occur. It may be that although each successive immunizing stratagem is intentional and introduced in the knowledge that the ideology is being altered in a slight respect, the whole sequence of immunizing

stratagems and their accumulated effect is unplanned and unforeseeable. An analogy with the evolution of language might clarify my point. Even if every change in the language were a conscious innovation, the total effect of all the unintentional ramifications of these intentional changes cannot be foreseen. No one living in Medieval England, for example, could have predicted the shape of today's English language.

If ideologists are indifferent to truth then why do they employ immunizing stratagems at all? There may well be cynical ideologists who have more dominant concerns than of truth, who are more interested in the perpetuation of their doctrine. But their audience is interested in truth. Perhaps the use of immunizing stratagems is an attempt to satisfy these conflicting interests. In any event, whatever the intentions of the propagandist, his audience coselects those elements that pass the filters of rationality (Percival 1994). The rationality of the propagandist's audience is one of Campbell's coselectors and part of the Popperian logic of the situation facing the propagandist.

2. The General Structure of Evasive Responses to Criticism

I maintain that the introduction of an immunizing stratagem will bring with it its own problems, which will need further immunizing moves. I also maintain that this process gets us further and further away from the original theory. If I am right and the succession of immunized theories are in fact different theories, how do we account for the appearance of continuity, for the relatedness of the theories? Once we grasp the general structure of the process that spurs the ideologist on from one theory to the next, we will have the answer to this question.

The general pattern of ideological evolution under criticism conforms to a schema proposed by Popper (1963) for the development of science: Problem 1 \sim > Theory 1 \sim > Error Elimination \sim > Problem 2

The original theory is an attempt to solve a problem. To some extent, then, theory production is designed. But exactly which theory will be produced is blind. Moreover, this conjectured solution often has unforeseen problems of its own, so we have problem 2. This new problem then prompts the modification (itself partly designed, partly blind) to the original theory to yield a different theory, theory 2. Popper argues that even the problems are theory-impregnated, and this is also

true, though not at first sight, of practical problems. Practical problems arise, Popper argues, because something has gone wrong because of an unexpected event. But this means that the organism has previously adjusted to its situation by some expectation, a pre-linguistic theory (1976, pp. 132-133). But we see that even the occasion for the first stage of Campbell's schema may involve a World 3 element.

To illustrate how this schema can be applied outside of what Popper would regard as science, we may point to the evolution of the idea of original sin. This example is taken from Wells (1988). The belief that God is just naturally leads to the expectation that the virtuous will be rewarded and the wicked punished (Theory 1). But the suffering of innocents makes it hard to believe that happiness and unhappiness are distributed according to this principle. Christians, therefore, had a problem in reconciling their belief with the world (Problem 1). Now the idea that God will compensate the innocent sufferer in heaven and punish the happy wicked in hell was unavailable to the early Hebrews because they had no belief in immortality. So the Christians supposed that the innocent sufferer was paying for the sins of some wicked ancestor. After all, it is always easy to imagine some wicked ancestor; any possible refuting evidence is more difficult to collect since one can hardly survey the whole of anyone's ancestry. The Christians were then armed with a new and "immunized" theory (Theory 2). But this in turn brought its own problem, since it implies that the good in every succeeding generation must be punished until the end of the world, and that there is nothing one can do about it (Problem 2). This then prompts the emergence of a revision in the earlier doctrine, an elimination of error. The Christian idea of Atonement is such a revision: we are cleansed of our inherited sins by the death of Jesus, provided we have faith in him (Theory 3).

The logic of the situation is often much more complex, as will be shown below in the analysis of how immunizing stratagems may lead to the break up of an ideological movement. The situation is perhaps better rendered with a branching structure in which each node represents an emerging faction dealing with the same problem in a different way (or with different problems, since factions may even disagree on what are the problems).

The above schema can help us to understand how Marx and Freud were led from one position to another in response to criticism. The schema helps us to see how the successive theories are, though different,

related to one another. The thread that seems to tie them together is a problem: specifically the sequence of unpredictable problems that the attempt to solve an original problem leads to. We also see that since the way criticism is dealt with cannot be predicted, any living doctrine must in one sense be a rambling structure. The rambling nature of the doctrine through time is no obstacle to our analysis, but its very object. (It is interesting to note at this point that over considerable time the importance of the various problems may shift considerably, either because later generations have forgotten the original primary problem or have different interests.)

The schema will also help us to identify immunizing stratagems. When identifying immunizing stratagems it is not sufficient to analyse individual statements. One has to relate the sequence of theories to the original problem that the first theory was meant to solve. For example, in the case of the swan hypothesis talk of essence could be identified as an immunizing stratagem if the original problem was to give empirical information about all swans - which in the hypothetical example is taken for granted.

3. Problems with the Demarcation Criterion and the Criticizability of Metaphysical Theories

Popper was from the beginning aware of several problems with his demarcation proposal, whose solution is very pertinent to the idea that ideologies such as Marxism and Freudianism are safe from empirical criticism. I argue that Marxism and Freudianism do not save themselves from empirical criticism by assuming metaphysical form, and that even in the absence of empirical criticism there is potential criticism from other metaphysical theories.

Popper realized as early as 1934, the year of the first edition of the *Logic of Scientific Discovery*, that a metaphysical idea can inspire the creation of an empirically testable theory. In that book he gave a number of examples, such as atomism (which inspired John Dalton's atomic theory explaining the regular proportions in which elements combine); the corpuscular theory of light (which inspired Planck's photon theory); and the theory of terrestrial motion (1934, p. 278). However, Popper notes that he was not fully alive to the fact that metaphysical ideas are

rationally arguable and in spite of being empirically irrefutable, criticizable, for example, as to their effectiveness at solving the problem for which they were proposed (1958, p. 206, footnote 2).

If it is accepted that what is important is the move from less to more informative theories, then interesting conclusions follow. For example, even if Marxism has been made into untestable metaphysics, it could be made testable again. Equally, Freudianism could be made testable. A Marxist or Freudian could be shown how their theories could be interpreted empirically and the theory promptly refuted. This need not be as arbitrary as it seems. Even lovers of metaphysics are constrained in their speculations by a whole network of what they regard as background knowledge (which may consist of both empirical and metaphysical theories) and their problem situation.

Popper also realized that there is a rational function to resistance to criticism; one can be too sensitive to criticism. A theory may need time to show its strengths: sticking to a theory even against very strong arguments. Moreover, it may require considerable debate to discover that what at first seemed purely metaphysical is actually empirical. The actual information and logical content of a theory is not only a conjectural matter, but is mostly unfathomable, a point I will take up later. The late physicist, Feynman, made a similar point when he stressed how difficult it is sometimes to work out how a new physical theory might be tested in the laboratory because it is often not even clear what, if any, empirical implications it has.

4. Empirical versus Metaphysical Criticism

It looks as though the ideologies most infamous for their apparent obstinacy in the face of criticism, take on a metaphysical form. Marx held that for all economies based on wage labour and a market in factors of production (i.e. capitalism) there is a tendency for monopolization of factors to increase and for an expansion and integration of workers' organisation. When monopolization had created one supreme world employer, the workers would take over its administration and institute communism. Apparently, Marx thought the revolution was imminent, certainly within his lifetime. The Marxist, however, can always say, it is often said, that communism will arrive eventually: the tendencies to

monopolization, he might protest, have been temporarily countered by opposing tendencies. Other utopian systems can escape direct refutation by making their prophecies apply to some eventual future rather than by putting a definite date on the coming of the new era. Can they be criticised in that form without first interpreting them empirically?

To clarify the logic of the sorts of systems we are talking about and the possible empirical criticism to which they could be put, let us take an example from chemistry. A classic metaphysical sentence is: gold has an acidic solvent. This is an irrefutable statement, for however far and wide one looks for such an acid without finding it, it is always possible to say that it exists at some other time or place. So is experience, our strongest critic, irrelevant to this type of statement? John Watkins (1958) has pointed out that experience can be brought in as a critic here indirectly via a well tested scientific theory which is directly testable. The metaphysical sentence in question is in fact incompatible with the well tested theory that gold has no acidic solvent.

But is such an analysis relevant to the Marxist's attempt to evade criticism? Yes, for like the spatio-temporally unrestricted singular statement about gold, the Marxist's apology is also a spatio-temporally unrestricted singular statement. Both would require a systematic search of the whole of space and time for a direct empirical refutation (or confirmation), which is obviously impossible. (Of course, the Marxist's assertion covers only future time, though it might be made to cover the past if he were desperate enough.).

If a Marxist did resort to this desperate manoeuvre, he would still be open to an indirect empirical refutation. Ludwig von Mises (1935) argued that without a price system, which communism would eliminate, there is no even equally adequate way to allocate resources. Against the desperate hope in the possibility of communism Mises pitted economic theory, a theory which makes many detailed empirical predictions.

One might argue that economics does not make predictions of the same empirical precision as does chemistry. But we can certainly say that economics has greater informative content than the Marxist's unrestricted singular prediction, and may still undermine the Marxist's case.

It is easy to assume that empirical observation is the strongest critic. The implication would be that if a network of ideas succeeds in shielding itself from empirical counter-evidence, it will have evaded, if not all sorts of criticism, at least the most damaging both psychologically and

logically. This may not be true. An interesting possibility is that perhaps opposing metaphysical theories are sometimes of greater weight than empirical observations. Watkins has shown how metaphysical theories serve to filter out some possible theories before they even enter the body of science; these theories do not even get discussed because they conflict with the prevalent metaphysical background assumptions.

Watkins' discussion of the influential role of metaphysical doctrines ('haunted universe doctrines') is highly suggestive in this context:

what informs and integrates the heterogeneous ideas of Augustine, or Bossuet, or Condorcet, or Burke, or Comte, or Marx is in each case a distinctive view of history which both shapes each of their interpretations of historical facts and suggests a certain kind of moral and political outlook....the moral-political suggestiveness of haunted universe doctrines indicates that large clashes of belief in the moral-political sphere need not have their origin in disagreement over moral principles or over observable facts. They may be generated, partly or wholly, by conflicting metaphysical interpretations of the world. (1958, p. 360)

There are other methods of criticism that can be applied to metaphysical theories. Galileo suggests a charming way to criticise doctrines that fail to exclude rivals by empirical test. Galileo was able to report that his telescope showed that the Moon was not a perfectly smooth sphere as the Aristotelians expected, but was instead marked by craters and mountains. One of Galileo's adversaries tried to defend the Aristotelian doctrine by suggesting that an invisible substance filled up the craters and covered the mountains so that the Moon was actually spherical. When Galileo asked him how the substance was detectable, he said it was undetectable. Galileo responded by saying that he was quite prepared to accept the hypothesis of the invisible substance, but insisted that it was in fact piled up high on the mountains of the Moon in such a way that the Moon was even more uneven than the telescope could reveal. Galileo's rejoinder allows one to see the inadequacy of the immunizing move, of making empirical testing irrelevant.

The same type of rebuttal can be applied to conspiratorial theories that have assumed an empirically untestable form. For example, suppose

some cynic asserts that all the set-backs in the workers' movement are instigated by undetectable groups of capitalists operating behind the scenes. One could counter this by saying that the set-backs are real and there are conspiratorial capitalist groups working against the workers' movement. However, their efforts are always unsuccessful, because they are always thwarted by undetectable renegade workers' groups who are the actual cause of the set-backs in the workers' movement. If the conspiratorial theory is successful on account of its lack of empirical testability, then the propagandist is prompted by the logic of his situation to try to counter the rival conspiratorial theory. But he can do this only by augmenting his theory with testable content.

We may conclude that even if an ideology assumes the form of a metaphysical doctrine it may yet be criticised, not only by unproblematically empirical theories, but also by scientifically acceptable metaphysical assumptions. The Marxist's retreat to unrestricted prediction does not save his position from criticism, but only creates other grounds for criticism.

5. Damaging versus Eliminating a Network of Ideas.

In correspondence the late W. W. Bartley III, partly conceding my point, argues that:

in a strict sense, the introduction of an immunizing stratagem may be tantamount to abandoning the position; but in practice it is more likely to be tantamount to damaging the position.
(February 13th, 1988.)

Jeremy Shearmur (in private correspondence) has made a similar criticism of my thesis. I think this is true, but misleading. Drastic revisions of a theory through the use of an immunizing stratagem are rare, for they are too obvious and unconvincing. The revisions are more often of a marginal nature.

Bartley's and Shearmur's disagreement with me rests on an unexamined assumption of theirs that there is a difference between modifying a network of ideas and making a new set of ideas, a form of essentialism that is false. One might say that a network of ideas may

evolve yet survive, in the sense that the fundamentals are retained, but one may then ask how fundamentals of a network of ideas would be defined other than as those elements of a modified network of ideas that are retained?

Even if we accept for the purposes of argument that to damage a position is not to eliminate it, the distinction breaks down when we look at the history of ideas. Metaphorically speaking, a sufficient number of injuries to a theory is equivalent to its death. Each intentional or unintentional concession made by an ideologue may be individually insignificant; but a sufficient number of insignificant differences makes a significant difference. Numerous, successive, slight modifications may lead from orthodoxy to radically different interpretations - to heresy. For example, how is it that the present day British Liberal party shares very little of the original Liberal party's doctrine as represented by Cobden and Bright?

In the light of such examples, when analysing the evolution of a network of ideas we ought to distinguish the following:

- (1) The uninterpreted terminology of the doctrine as embodied in books etc;
- (2) The actual interpretation placed on the terminology;
- (3) The interpreter's theory about how his interpretation compares with previous interpretations (his own and others').

(If we wish to include ceremony etc, we can substitute "symbolism" for "terminology". The notion of an uninterpreted term is purely conceptual; in nature perhaps everything we attend too receives some kind of interpretation.)

When Campbell (1979) said "All self-perpetuating belief communities are tradition-ridden, viewing current events through the spectacles of their past", he was right about their *intention*. However, It is important to recognise that (1) and (3) may remain constant while (2) changes quite dramatically. For example, in the simple case discussed, the words "all swans are white." are retained, but the interpretation placed on them is altered considerably. We can also imagine that the person who proposed the claim about swans thinks that his later interpretation of his statement is exactly the same as his earlier interpretation - when challenged he might retort: I thought that all along.

6. *Do all Immunizing Stratagems Abandon the Theory for whose Protection they were Introduced?*

It is not my aim to show that criticism cannot be deflected in any way by logical means. My aim is to show that many so-called immunizing stratagems actually abandon the theory they were introduced to save, while many others often lower the chances that the theory in question will be reproduced and successfully compete with other theories. But to take account of Bartley's and Shearmur's criticism I need to distinguish more precisely between a privileged subset T of a set of assumptions, and a useful but dispensable subset A. The more subtle claim then is that by tinkering with the subset of dispensable assumptions A, any T may be preserved in the face of any counterevidence.

Suppose T & A yields as a consequence the implication e, but the accepted counterargument implies -e. If the response of an ideology to criticism is to modify its assumptions then it may replace A by A' in one of three ways:

- (1) T & A' implies e, where A' = -e.
- (2) T & A' implies -e, where -e is not derivable from either T or A' alone.
- (3) Such that neither T & A' implies -e, nor T & A' implies e.

If (1) then information content will be lowered and each successive theory will become increasingly a burdensome hotchpotch of unrelated hypotheses, sacrificing by incremental steps the preference for systematic organisation. The system also becomes more difficult to learn and pass on. Moreover, there is no proof that a replacement A' that is consistent with T can always be found (T and -e may be inconsistent).

If (2) then T is retained and also used systematically in the derivation of -e. There may even be an increase of information content. But this latter would make T & A' even more open to criticism. Again, there is no general proof that for any counterevidence -e against any theory T & A there is always a suitable A' that in conjunction with T will yield -e.

If (3) then there is clearly a loss of information content. Weakening A so that T & A' no longer implies e may also necessitate a loss of other implications that were important in solving problems for which T & A was initially adopted.

As I have indicated, not all immunizing stratagems involve modification to the information content of a theory's assumptions. Some

that at first do not seem to fall into this class can be interpreted this way, but not all. It would be helpful if I made a list of the types of immunizing stratagems and then examine which ones involve the abandonment of the original theory, and which impair the theory's chances of spreading.

- (1) Denying the refuting evidence, e.
- (2) Reinterpreting the theory as a definition or the implication of a definition.
- (3) Adding other assumptions to T in the presence of which the resulting theory is consistent with or implies e.
- (4) Subtracting assumptions from T such that the remaining set of assumptions is consistent with or implies e.
- (5) Reinterpreting the theory as essentialistic.
- (6) Introducing the idea that the theory is beyond the capacity of human reason to criticize or test (e.g., God moves in mysterious ways).
- (7) Introducing ad hoc exclusion clauses to T for special cases.
- (8) Reorganising the conceptual structure of the theory.

(1) An example of the denial of evidence is Marx's attitude to the price of goods offered for sale that are not mass produced commodities, such as honour, conscience or unworked land etc. In these cases, Marx asserts, the prices are imaginary, like certain quantities in mathematics (1867, p. 105). In saying this Marx expresses his confusion. It may look as if he is saving the labour theory of value but he is substituting another theory instead, one that may well be implied by the original theory but certainly one of much lower information content.

Suppose the refuting evidence, e, is denied under all circumstances. For example, in the swan case, the person who advanced the theory that all swans are white may simply deny that any black swan presented to him is black.

Prima facie, this does not look like a case in which the original theory is abandoned. But let us look more closely. At least some of the information content of an empirical statement is logically equivalent to the class of basic statements with which (perhaps in the presence of other assumptions) it is inconsistent. In other words, the basic statement that constitutes e would be part of the meaning of T. Now if no basic statement is treated as inconsistent with a purported empirical statement, then we may infer that it is, after all, non empirical. As a corollary, it follows that the original claim was either wrongly presented as empirical, or was empirical and was later abandoned for another theory with the

same terminology. In either case, the original claim has been abandoned.

(2) Reinterpreting a theory is in some cases abandoning the earlier theory; in some cases it is simply changing the conceptual structure without changing the theory. We will see in the case of Marx's Labour Theory of Value how the crucial term "socially necessary labour time" is reinterpreted several times, the total amounting to an abandonment of both the original theory and the original problem.

The assumption that one can modify a theory without abandoning it does have some truth to it. One can completely reorganise the conceptual structure of a theory without changing its empirical content. Popper himself has been keen to make this distinction between a theory and the concepts in which it is expressed (1982 p. 42). The same theory may be formulated in many different ways and may use different conceptual schemes.

Changes in the conceptual system employed by a theory may function as protection against criticism, since it may disarm the critic - it may appear to the critic that the theory has been abandoned under the pressure of his criticism, whereas in fact the old theory is retained under the (intended or unintended) camouflage of the new concepts. However, such an effect has costs for the ideology's survival value that may be overlooked: (a) the ideology has to be relearned - a transmission cost; (b) to the extent that the change of concepts is unintentional there is a loss of understanding of the theory.

Can a propagandist guarantee that by introducing ad hoc purely abbreviative definitions to evade criticism that the system will not incur new unpredictable commitments that are themselves open to criticism? One might think that a purely abbreviative definition adopted as camouflage would be neutral, but as Popper argues, some abbreviative definitions are creative in the sense that they alter what can be derived from the theory (1982b, p. 170). A definition is creative if there are theorems not containing the defined term that cannot be derived without the help of the definition of the term. There is no routine way of telling whether a definition is creative or not, so even seemingly trivial evasive redefinitions may have unwanted but unforeseeable repercussions in the rest of the system, perhaps creating other more serious avenues for criticism. Moreover, even if an evasive definition is at first purely abbreviative, it may become creative by the removal or addition of an axiom to the system.

(3) Adding assumptions. One might at first think that there are two main ways in which a theory T may be immunized through changes in the assumptions of the theory: (i) a move from T to T' (where T' is the conjunction of T and one or more auxiliary assumptions, denoted by B); (ii) a move from T to X (where X is T minus some of its assumptions, perhaps with replacements). Only (ii) represents the abandonment of assumptions of T, and its replacement by another theory. One might argue that (i) preserves the original theory within the substitute, and therefore immunization can preserve an ideology. Thus Lakatos says:

For instance, we may have a conjecture, have it refuted and then rescued by an auxiliary hypothesis which is not ad hoc in the senses which we have earlier discussed. It may predict novel facts some of which may even be corroborated. (1970, p. 175)

But (i) is not a logically possible immunization. The modified theory cannot be consistent with the falsifying evidence if one simply adds extra assumptions that increase information content. For suppose theory T is false with respect to evidence e; then, since a conjunction is false if and only if one of its conjuncts is false, any conjunction consisting of T and an extra assumption B will also be false with respect to e.

This general point can be applied to Boudon's treatment (1986). Boudon's analysis suffers from a lack of logical sensitivity, which grossly misleads him. His failure to distinguish between different components of a theory allows him to infer that a refuted theory can be consistently retained by adding extra assumptions:

Suppose that a physicist of Newton's time discovers that a planet is deviating from the orbit assigned to it by theory T. T could nevertheless be kept thanks to an adventitious hypothesis. (1986, p. 161)

and because of exit costs of leaving T for T' (an alternative theory) and entry costs (learning a new language etc.) of adopting T',

people will try to keep T going by trying to reduce the inconsistencies between T and the facts of the real world by means of adventitious hypotheses. (1986, p. 162)

This argument amounts to a simplification of Lakatos's argument, discussed below. As we will see, Lakatos makes a careful distinction between different components of the theory at issue.

To make this point clearer consider the case of Leverier and Adams. They did not reject Newton's laws of motion and gravity. Newton's theory, consisting of the laws of motion and of gravity conjoined with auxiliary assumptions regarding the number, mass, position and acceleration of the planets and the Sun, was inconsistent with the observation reports of the motion of Uranus. Leverier and Adams introduced another assumption: the existence of the planet Neptune, with a certain mass, position and acceleration. Now if an ideology adopted this tactic it would be adding to knowledge and sustaining itself. But it must be borne in mind that the augmented theory is now more open to criticism, so is hardly being guaranteed from it.

But the above is not accurate enough. What Adams and Leverier did was to deny one of the auxiliary assumptions of Newton's theory: that there were no other planets in the solar system but Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus. Therefore, the modified theory of Adams and Leverier actually contradicted Newton's theory (though not the laws of motion and gravity alone). If any ideologist did this he would be abandoning his ideology.

(4) Subtracting assumptions that reduce the information content of an ideology effectively means that the original theory is abandoned. Though subtracting disjuncts may increase information content, for a statement p is logically stronger than p or q . We will see below that Lakatos effectively replaces a conjunction of premises in Newton's theory by their disjunction and thereby empties the theory of much content.

(5) Reinterpreting the theory as essentialistic. As we will see in the case of Kepler, if he had tried to sustain his masters' position on the circularity of planetary orbits by asserting that the orbits of planets are essentially circular, he would have replaced a hypothesis with much content with one of possibly zero content. But not all essentialistic hypotheses are completely devoid of content. Hume says of the peripetetics that when they were asked for the cause of a phenomenon, they would resort to faculties or occult qualities. They would say that bread nourishes by virtue of its nutritive faculty and senna purged by virtue of its purgative powers (1779, p. 73). Hume took these hypotheses as devoid of content, but in fact they could be interpreted so that they

rule out some possibilities, such as the class of causes which lie outside the bread or senna. But as with the swan hypothesis, in making the interpretation, one would have to check it against the proposers' background knowledge and problem situation. However, if the theory is empirical in Popper's sense, then replacing it with an essentialistic theory will abandon much content. Essential explanations often imply an ultimate explanation. Essentialist immunizations run the risk of offending the desire for more information because they rule out further generalisations, explanations of greater depth. It may then lose in a competitive struggle with other ideas that address the same problems.

(6) This sort of tactic is very interesting. Neither Marx nor Freud resorted to it as it would have been completely anathema to their enlightenment inclinations, but it is a common practice in religious circles. It is an example of what Bartley would call a retreat to commitment. But its strength can easily be exaggerated. To function properly it must be kept under control, for it may backfire. For example, a sceptic may retort: if God moves in mysterious ways, how do you know that it is God and not the Devil that speaks to you on any given occasion? God's command to Abraham that he sacrifice his son Isaac was, I suppose, a mysterious way of acting. But when Abraham obeyed God's command he did not first try to test the identity of the voice that spoke to him. But why not? It could have been the devil, as far as he knew - if, as you say, God moves in mysterious ways. So there is a counterargument. But the original theory - that God exists - looks as though it has been retained: has it? Well, at first we have the confident assertion that a unique being answering to a definite description exists (i.e. all powerful, completely benign). Whatever else may be true of this being, it does follow necessarily that if he did exist there could be no evil for he would not suffer evil to exist. But in the face of counter-evidence this implication is denied. In so far as it is denied we have a different theory before us. The doctrine that there is a completely unfathomable mysterious something seems to be almost no doctrine at all. How can we check that there is not more than one unfathomable God? The doctrine of original sin, which attributes the evil in the world to human free will, forgets the suffering of innocent new-born children obviously too young to know of good or evil.

(7) The introduction of ad hoc exclusion clauses is also a case of abandoning the original claim. For example, suppose someone advances

the theory that bread nourishes, but then notices that a certain batch of bread kills some people. If he then says bread nourishes, except that particular batch which killed those people, he has reduced the content of his claim and therefore abandoned the original theory. As more counterinstances are dealt with in this way the theory becomes increasingly a hotch potch of unrelated hypotheses, losing its systematic character. It not only becomes clumsy in application but more difficult to learn and pass on.

(8) The propagandist may alternate between two or more theories. This is an interesting case in which the original theory is not completely abandoned. It is quite possible that two interpretations of the text are maintained, each being brought to the fore when powerful criticism makes it difficult to assert the other. Frank Cioffi has noted this phenomenon in connection with Freudianism:

It is characteristic of a pseudo-science that the hypotheses which comprise it stand in an asymmetrical relation to the expectations they generate, being permitted to guide them and be vindicated by their fulfilment but not to be discredited by their disappointment. One way in which it achieves this is by contriving to have these hypotheses understood in a narrow and determinant sense before the event but a broader and hazier one after it on those occasions on which they are not borne out. Such hypotheses thus live a double life - a subdued and restrained one in the vicinity of counter-observations and another less inhibited and more exuberant one when remote from them. (1970, p. 474)

The bold version is still prized for its richness of information content and so is brought forward in certain circumstances.

How does this phenomenon fit in to the evolution of an ideology? We may conjecture that this is a typical stage in the response of an ideology to powerful criticism. First we have the pristine doctrine promulgated faithfully with great confidence. Then, in response to criticism, we have the original doctrine supplemented by the immunized version, brought forward in appropriate circumstances. Most commentators have overlooked the increased burden of the excess theoretical baggage that this alternation involves: new converts have to learn not only the original

theory (usually quite cumbersome in itself) but also the adapted one. The likely consequences are (a) increased errors of transmission and (b) simple confusion, neither of which contribute to the morale of the movement and may impair the propagation of the ideology.

It is my guess that this stage tends to be followed by one in which the original doctrine is completely supplanted by the adapted version. Thus we have:

- (1) Original doctrine;
- (2) Original doctrine plus adapted doctrine;
- (3) Adapted doctrine.

To further explore Bartley's suggestion that the use of immunizing stratagems may only amount to a modification and not an abandonment of the theory being "protected", I must also examine Lakatos's distinction between a hard-core and protective belt and Duhem's problem, since both of these seem at first sight to show that by tinkering with a system of hypotheses, refutation may be avoided and therefore a privileged sector of an ideology may be retained regardless of the facts. These ideas make more precise the suggestion that the "essentials" of a system might be guaranteed from criticism and perpetuated even though the system evolves in response to criticism.

I intend to argue that all I need to sustain my thesis that no system of ideas can be guaranteed from criticism is no more than Bartley and Shearmur concede: that the immunizing stratagem may simply "damage the position".

7. Hard Core versus Protective Belt

Lakatos makes the distinction between the hard-core of a theory, which is preserved in the face of unfavourable evidence, and a protective belt of hypotheses which may be changed to accommodate any unfavourable evidence. Lakatos argues therefore that no core scientific theory forbids any observable state of affairs. If this were true an ideology could in principle adopt this kind of stratagem to deflect criticism from a privileged portion of its structure. I argue that not only is this not always logically possible, but even if it were it assumes superhuman powers of memory and reasoning. First, most commentators have been too quick to assume that finding a suitable change in the protective belt is easy:

creating such a protection may require more time, effort and genius than creating an alternative core theory. Second, any attempt to guarantee that the changed protective belt will not adversely affect the hard-core is doomed to failure: the line drawn between hard-core and protective belt is a conjecture, and one cannot always rule out the possibility that some remote logical consequence of the change to the protective belt will not be in conflict with the core theory. The same points apply also to Duhem's argument.

We can explain Lakatos' distinction between the hard core and the protective belt with the help of a story about an imaginary series of problems in Newton's research programme (1970, pp. 100-101). A Newtonian using Newton's mechanics, law of gravitation plus generally accepted initial conditions, calculates the trajectory of a newly discovered planet. However, the planet deviates from the calculated trajectory. The question then is: Does our Newtonian place the blame on Newton's theory? No, he attributes the failure in prediction to his statement of the initial conditions: there is an as yet unobserved planet p' which perturbs the trajectory of p . The Newtonian calculates the mass, orbit, etc. of this planet p' and asks an astronomer to try and detect it. Even if the astronomer fails to observe it, the Newtonian may sustain his allegiance to Newton's theory by conjecturing that the planet p' is too small to be observed even with the most powerful of current telescopes. Lakatos' point is that with resolution and enough ingenuity, the Newtonian can select a part of his set of accepted statements as a privileged sector to be made safe from criticism by appropriate changes in other beliefs.

Lakatos drew the conclusion that in fact no scientific theory can be refuted because the theorist can always introduce auxiliary hypotheses to deflect criticism from the theory:

exactly the most admired scientific theories simply fail to forbid any observable state of affairs. (1970, p. 100)

However, Popper has undermined this contention. Lakatos promises to back his "characteristic story" up with a general argument, but this general argument can only succeed if the most admired theories are denuded of part of their fundamental content. In assessing Lakatos' argument, Popper says, it is important to be clear that his thesis does not depend on arguing from defective observations:

even if there were a firmly established empirical basis to serve as a launching pad for the arrow of the modus tollens: the prime target remains hopelessly elusive. (1970, p. 100)

Lakatos feels that he can generalize from the "characteristic story" to the most respected theories, such as Newton's. But, Popper argues, this would require the assumption that any deviations of a planetary orbit from its predicted path can be accounted for by Newton's theory by postulating the influence of some other planet (more generally, massive body). However, as Popper points out, there are an infinite number of planetary orbits which cannot be accounted for in this manner. (To be precise, there is an infinite set of measure 1 such orbits.) For example, Newtonian gravitational theory cannot explain a square or a triangular orbit, no matter what is assumed about the mass, position, etc. of other planets. Lakatos's story, Popper concludes, cannot therefore be characteristic, but is in fact quite exceptional.

Lakatos asserts that

some theories forbid an event occurring in some specified finite spatio-temporal region (or briefly, a "singular event") only on the condition that no other factor ... has any influence on it. (1970, p. 101)

Putnam argued the same point in Popper (1974, p. 221). Lakatos draws from this the conclusion that such theories never alone contradict a basic statement, but at minimum (he says maximum, but the meaning is clear) the conjunction of a basic statement with a universal non-existence statement saying that no other factor is at work. What Lakatos overlooks is that his result is obtained at the price of emptying Newtonian theory of important content. Popper argues that Newton's theory of gravitation amounts to the thesis that all bodies in interplanetary space not only move according to Newtonian dynamics, but their movements can be explained by an appeal to *gravitational forces alone* (1974, p. 1008). If this is true, then Newton's theory does not allow the possibility that other factors may be at work, as Lakatos's *ceteris paribus* clause suggests, but actually denies their operation, making the theory much stronger logically. And this is why Newton's theory was refuted by the first rocket that travelled

outside the earth's atmosphere.

O'Hear makes the same mistake as Lakatos. O'Hear criticizes Popper's counterexamples to Lakatos' thesis, saying that the peculiar orbits of the planets could be produced by powerful rockets on the planets involved. We will ignore for the moment that this state of affairs is ruled out by Newton's gravitational theory for the reason I have just discussed, to see how strong O'Hear's argument is in other respects. It is not clear that such an arrangement could produce rectangular orbits, and O'Hear supplies no argument here. O'Hear finds it sufficient to say that Lakatos's general point is grasped: "that such explanations are always possible..." (1980, p. 102). But O'Hear seems not to have taken heed of Popper's reply to Lakatos in the Schilpp volume, where he points out that Lakatos provides no general argument for such possibilities: it is far from obvious that such explanations are always possible.

Watkins (1984) agrees with and elaborates O'Hear's argument, making use of his notion of observational predicates. Watkins begins with a useful distinction between the "fundamental" assumptions of a scientific theory and "subsidiary" assumptions, which Popper calls the initial conditions. In Watkins' account the "fundamental" assumptions of a scientific theory are universal statements. In the case of Newton's theory the fundamental assumptions are the law of gravitation and the laws of motion. The subsidiary assumptions would be, for example, statements concerning the position, mass, number and acceleration of the planets. Watkins calls the fundamental assumptions taken alone the "core theory" T, and the combination of this with subsidiary assumptions the "fleshed out theory", T & A (p. 324). If we make this distinction, Watkins argues, then we may say, along with Lakatos, that all core theories fail "to forbid any observable state of affairs". This, Watkins says, is because the "core theory lacks the observational predicates needed for a possible conflict with observation reports" (p. 325). Only the subsidiary assumptions can supply these predicates. He infers from this that Popper's proposed examples of potential falsifiers of Newton's core theory do not count as such.

Paraphrasing O'Hear's conclusion Watkins writes:

Newton's laws of motion plus his law of gravitation say nothing about the physical makeup of the planets; in particular they do not rule out the possibility that the planets are enormous

rocketlike devices that can accelerate themselves in all sorts of ways. (p. 326)

In a straightforward sense, it is true that Newton's core theory says nothing about the chemical constitution or size or mass or structure of the planets. However, the laws of motion rule out an infinite number of logically possible accelerated motions of objects with mass, and therefore motions of planets or rockets. According to Newton's core theory, therefore, rockets cannot "accelerate in all sorts of ways". Newton's core theory may not have the predicates "rocket" or "planet" but it certainly has the predicates "acceleration" and "mass", and all one needs to know about the planets is that they have mass for the core theory to rule out infinitely large classes of their possible motion. One could say, for example, that the second law of motion rules out the possibility of masses moving in accord with the law $F=ma^2$. This is a little unconvincing because in order for these laws to contradict one another one has to assume that bodies with mass exist.

But I have a more convincing argument. A rocket cannot accelerate from zero to any finite velocity instantaneously. To modify Popper's rectangular orbit example, we can imagine a rocket moving at constant velocity v along each of the borders of the rectangle, stopping instantaneously at each corner, remaining stationary for an hour, then moving off instantaneously with constant velocity v to the next corner. (v could be any one of an infinite number of finite velocities.) We can imagine this, but according to Newton's second law of motion alone it is impossible. Since force equals mass times acceleration an instantaneous change in velocity would require an infinite force. Could an obstinate Newtonian just postulate the existence of infinite forces? No, for that would make the mass of the rocket indeterminate because an infinite quantity divided by another infinite quantity is indeterminate. (For example, aleph zero divided by aleph zero can have any value from 1 to aleph zero.) But the meaning of Newton's law is that given any two of the values, F , m , a , the equation will yield a determinate answer for the third. Therefore, Newton's second law of motion taken alone rules out infinite accelerations.

In arguing against O'Hear and Watkins here I have allowed their argument considerable latitude and still found it wanting. I ignored the fact that the core of Newton's system contained the assumption that all

the forces acting on the planets were gravitational. But this assumption is implicit in the way Newtonians solved their problems. It effectively rules out O'Hear's rocket propelled planets; a report of such phenomena would constitute a falsification of Newton's core theory and so could hardly serve to protect it. Watkin's should note here that Newton's core theory does not have to mention rocket propulsion for it to deny by implication rocket propulsion of the planets, since such propulsion implies that planetary motion is governed (at least partly) by non-gravitational forces.

8. *Duhem's Problem*

Duhem's problem is the problem of attributing the failure of a prediction. Suppose someone wants to test a theoretical statement B_1 . If a set of assumptions B_2, \dots, B_n , are required in conjunction with B_1 to deduce a prediction g , and the result of the experiment, e , contradicts g , one cannot conclude that B_1 must be false. One can, however, deduce the falsity of the conjunction $B_1 \& B_2 \dots \& B_n$. Therefore, Duhem concluded:

the physicist can never subject an isolated hypothesis to experimental test, but only a whole group of hypotheses. (1914, p. 187)

Even if we know that exactly one hypothesis is false, no experimental outcome will enable us to attribute the fault exclusively to one of the hypotheses.

The relevance of Duhem's argument to ideology is this. The ideologist may seek to protect a privileged part of his system of assumptions in response to empirical criticism by jettisoning those assumptions he regards as of little importance. This privileged part of the ideology would then be guaranteed against empirical criticism. Popper tries to show how in some circumstances such a defence would be ruled out because we could focus the criticism onto just one hypothesis by comparing two systems:

Admittedly, Duhem is right when he says that we can test only huge and complex theoretical systems rather than isolated

hypotheses; but if we test two such systems which differ in one hypothesis only, and if we can design experiments which refute the first system while leaving the second very well corroborated, then we may be on reasonably safe ground even if we attribute the failure of the first system to that one hypothesis in which it differs from the other. (1957, p. 132, footnote)

If this were valid it would also further undermine Lakatos's argument, for the crucial experiment may focus on what he calls the hard-core.

However, Watkins has shown Popper's argument to be invalid. Watkins begins by paraphrasing Popper's description of the two systems to be compared by a crucial experiment:

denote that one hypothesis by B_1 and the large number of hypotheses common to both systems by B_2 ; and let B_1' be the hypothesis which replaces B_1 . (1984, p. 321)

Watkins compares theories to recipes. Suppose we have two recipes for a pudding, one uses cinnamon P and the other P' uses nutmeg instead of cinnamon. P' proves to be a better pudding. But this does not mean that nutmeg is gastronomically superior to cinnamon. Perhaps by keeping cinnamon and varying other ingredients in P the chef might produce a pudding even better than P' . Something analogous holds for theories:

Perhaps B_1 is true and B_1' is false, but B_1' is the better partner for B_2 because there is an error in B_2 that is cancelled out by a compensating error in B_1' . (p. 322)

One may quibble with the assertion that the errors in the assumptions are "cancelled out", for the falsity content of B_2 & B_1' is no less than (and possibly greater) the sum of the falsity contents of B_2 , B_1' taken separately. Watkins' point may be stated more accurately: a conjunction of two or more false assumptions can yield true deductive consequences that none of the assumptions taken separately could yield.

An example in political theory would be the following:

B_2 = Communism will emerge in a society if and only if more than 70% of workers in that society are employed in industry and involved in unions.

B_1' = Russian society in 1987 had less than 20% of its workers employed in industry and involved in unions.

(Suppose B_1 = Russian society in 1987 had 71% of its workers employed in industry and involved in unions.).

Even though both B_2 and B_1' may be false, together they imply the true statement that communism did not emerge in Russian society in 1987. Whereas if the true statement B_1 is conjoined with B_2 we may deduce the false statement that communism emerged in Russian society in 1987.

I freely accept that it is sometimes possible for an ideologist to protect a particular sector of his assumptions from some counterevidence that undermines his assumptions taken as a whole by tinkering with what he regards as trivial auxiliary assumptions. But no one has yet proven that this can always be done for any particular system for any counterevidence. It has yet to be shown, therefore, that an ideology can always in principle be guaranteed from criticism on account of Duhem's thesis.

There is also a purely practical problem for the propagandist in protecting his privileged sector of assumptions. Propagandists have a limited reservoir of immunising stratagems, especially in the short run, and persistent criticism will tax the most inventive apologist. Alan Musgrave makes a similar point in regard to theorists in science. Arguing against Lakatos's idea that scientists can always defend the hard core of their research programme, Musgrave points out that outstanding Newtonians tried for fifty years to explain Mercury's perihelion without having to abandon Newton's laws, but despite their undoubted ingenuity they failed (1978, p. 195). Moreover, each movement faces competition for adherents from many other movements; each has more critics than defenders. The "protective belt" may then collapse.

9. Changing Demarcation between the Hard Core and the Protective Belt

Again, in assessing the relevance of Duhem's and Lakatos's idea to ideological survival, one must take a long-term view. Who is going to police the distinction between hard core and protective belt of an ideology down the centuries? As I pointed out before, the terminology of a system of ideas needs interpreting, so it is a conjectural matter as to whether the system of ideas is being reproduced or not. Moreover, later adherents

may well disagree, wittingly or unwittingly, with earlier adherents about what constitutes the privileged sector of beliefs, especially when the earlier adherents are no longer around to argue the point. These disagreements may be genuine mistakes in interpreting the work of their predecessors.

Marxism is a particularly good example here. Marx had a very definite idea of what communism was: an industrially advanced society much more productive than our own without the buying and selling of factors of production. There would definitely be no market in factors of production: this was regarded as the indispensable part of Marxism. In the 1920s and 1930s a devastating attack on the possibility of communism in this sense was launched by Ludwig von Mises and his pupil Friedrich von Hayek (Mises 1935, pp. 87-130).²

It took time for their arguments to sink in, but by the 1950s and 1960s so-called Marxists were advocating "market socialism". There had been no explicit acknowledgement of error, but the old message had been dropped by many. Not many Marxists today are aware of the fact that they are no longer reproducing what Marx said. Many self-styled Marxists are in fact espousing some form of pre-Marxist socialism; they are Owenites, Proudhonists, etc, but rarely non-factor-market society Marxists. Many new forms of "Marxism" have also emerged, further threatening the original demarcation between the "hard core" and the "protective belt".

I would like to develop a general argument to show that there are limits to the ability of a propagandist to defend that part of an ideology that Lakatos might call a "hard core" and Watkins a "core theory". The logical ramifications (i.e., the information content and logical content) of a theory cannot be fully surveyed. Therefore, when modifications are made to the protective belt or subsidiary assumptions A, the theorist cannot always conduct a consistency proof to ensure that remoter consequences of the changes in A plus other assumptions remote from the core theory will be consistent with the core theory.

In the following argument I will adopt Watkins' terminology and distinctions with additions, but the same argument would carry through using Lakatos'.

T = The core theory of the ideology.

A = The subsidiary assumptions of the ideology.

A' = Modified A.

W = Total world view. The set of all assumptions, implied and asserted, that the individual maintains either in belief or in argument.

$b = A \setminus A'$ (that which is in A' but not in A).

Assume that $T \& A$ implies e , but a counterexample c , which implies not e , is responded to by the replacement of A by A' , which in conjunction with T implies not e (or is at least consistent with c). Assume that the modification of the subsidiary assumptions A to make A' amounts to the assumption b . Now, it is quite possible for b to be consistent with $W \setminus T \& A'$ and consistent with T , but for $W \& b$ to be inconsistent with T . For the propagandist to guarantee, therefore, that the adoption of A' would save T , he would have to survey the whole of his world view, which is at least practically impossible. I have assumed that the world view W of the ideologist includes and is larger than his ideology $T \& A$. Both Gellner and Shils have pointed to the fact that the ideologist exists in a surrounding culture that they cannot fully divest themselves of (Shils 1968, p. 67). This seems to be an inescapable part of the logic of the propagandist's situation: even if one assumed quite unrealistically that the propagandist himself was but a cipher of his ideology, his audience and converts have a much broader and richer belief system that may interact in unforeseeable ways with the ideology.

But the difficulty is even worse for the real life propagandist. Ideologies are adopted by people partly because they provide explanations or interpretations of new and unforeseen events and developments. These may include particular events - coup d'états, wars, economic slumps etc - or theoretical developments. To do this an ideology has to adopt changes in the world view of its adherents and new subsidiary assumptions A' to interpret or explain these. Even granting the propagandist superhuman powers that enable him to eventually perform a consistency proof for each modification it may be some time before the contradiction comes to light. By that time the assumption b may then have acquired great importance for explanatory and/or rhetorical reasons in maintaining adherence to the ideology. Costs in terms of learning alternative interpretations without b may also be considerable. Some of the content of T may then be sacrificed in order to retain b .

The history of Marxism supplies an illustration. Enrico Barone (1908) developed a system of simultaneous equations that described in formal terms the structure of input/output functions and prices in an advanced industrial economy. Many Marxists who had been stung by

Mises' argument mistakenly welcomed his paper as a vindication of their hope that an advanced industrial economy might be run according to a single plan. But Barone's paper actually helps one to understand how complex the problem of economic calculation is and to better understand Mises' argument. One could argue that having learned Barone's paper with the intention of bolstering their position Marxists became as an unintended consequence more open to counterargument. This argument takes Watkins' distinction between the core theory and the subsidiary assumptions for granted. It then shows that even with this clear distinction between the part of the theory to be preserved and the expendable part, there is no guarantee of maintaining doctrinal integrity and propagational success. But with ideologies such as Marxism and Freudianism, there is no clear distinction between the core theory and the subsidiary assumptions. Thus these ideologies are even more open to such self-destructive developments than would appear at first sight in the light of a straightforward application of the analyses of Watkins, Duhem and Lakatos to ideologies.

10: Factionalism Generated by Unpredictable Emergence of Incompatible Immunizing Stratagems

There is a tendency for different members of the original group to favour different stratagems, with the typical rise of conflicting factions which battle it out between them, often with a tenacity and vehemence worthy of a family feud. Indeed, the intensity with which factions squabble among themselves is greater than their quarrel with incompatible but non-heretical groups. Factions hate most those heretics most close to them. Thus a Stalinist has more venom for a Trotskyist than he has for a Classical Liberal. These factions in turn may split under the intensified criticism. The greater concern with close heretics than with distant opponents is a tacit acknowledgment that marginal deviations, if not checked, can eventually add up to great schisms. Hence the frequent resort that Christianity has made to special Councils to lay down explicitly what is to count as dogma. This is a very definite cost to the employment of immunizing stratagems. With the formation of such factions the original demarcation between the "hard core" and the "protective belt" can easily become blurred and abandoned. In any living

ideology there is a continual struggle between the attempts to achieve conformity and the unintended deviations tending to the formation of factions. It is a form of unstable equilibrium in two senses. Firstly, even in the most stable reproduction of the ideology, there is a continual oscillation between deviation and correction. Secondly, once a faction is formed the forces leading to deviation increase dramatically. The best analogy in mechanics is a balancing act.

11. *Unfathomable Implications of an Ideology*

The question naturally arises: could not some very determined propagandist settle the problem as to what stratagems will be needed and used in advance and so keep the faithful on the one true path? All new recruits could be specifically enjoined to keep to these and only these stratagems. This would be analogous to Lakatos's "positive heuristic". However, this problem of propaganda is in principle unsolvable.

This circumstance springs from certain logical properties of theories which make it impossible for any individual or group to foresee what specific immunising stratagems will be needed in response to awkward questions and criticism. The work of Church and others can be used to show that no ideologist could construct such a proven complete set of immunizing stratagems.

What our propagandist needs in order to guarantee his position in advance is an effective method of listing all and only the possible counterexamples to his ideology, so that he can check whether any proposed set of immunizing responses would meet all these possible difficulties. An effective method is one that can in principle be carried out by a machine: at any stage the method unequivocally determines how the computation shall proceed and terminate. According to Church's theorem of the undecidability of the predicate calculus this cannot be done. Imagine an arbitrary set of sentences constructed according to the rules of the predicate calculus. Church's theorem amounts to saying that there is no mechanical way of sorting these sentences into two sets: the set consisting of those sentences that are tautologous consequences and the set of those sentences that are not tautologous consequences. Because although any tautologous theorem will eventually be placed in the tautologous set, there is no way of telling of any sentence not so placed

whether it is non-tautologous or whether the method has yet to class it as such. A consequence of Church's theorem is that any theory with universal and existential quantification ("all" and "some" statements) plus unambiguous cross-referencing cannot be supplied with an effective negative proof of theoremhood. In the propositional calculus the truth-table method can determine eventually whether any particular sentence is a logical truth or not. The truth-table method is an effective positive and negative test of logical truth for this system. However, only a positive test is available for the predicate calculus. But the predicate calculus describes the formal structure of the most interesting part of ideologies: the claims to universal significance.

Now, every non-tautologous theorem of a theory is false in some interpretation; i.e. every non-tautologous consequence has a possible counterexample. It follows that there cannot be an effective method of constructing counterexamples, for if there were then there would be an effective method of determining of any sentence whether it is not a tautologous theorem, contrary to Church's thesis. Our propagandist cannot therefore determine in advance a set of immunizing stratagems that would deal with all and only the possible counterexamples to his ideology because he cannot even determine the set of possible counterexamples to check them against. Of course, the argument does not exclude the possibility of the propagandists' simply guessing correctly what the possible counterexamples to his ideology are. However, we will see below that even in principle such counterexamples cannot be listed: they are indenumerably infinite.

The non-trivial implications of any theory are infinite. No individual or group could therefore survey all possible criticisms and prepare standard responses to deal with them. To develop this argument it is helpful to distinguish between two associated but different senses of the content of a statement or theory, which Popper has called 'logical content' and 'informative content'.

The logical content of a theory consists of the set of all (nontautological) consequences which can be derived from the statement of the theory. The information content of a theory consists of the set of all those statements which are logically incompatible with the theory. The latter idea derives from the intuitive idea that a theory tells us more the more it prohibits or excludes.

There is a one-to-one correspondence between the information

content and the logical content of a theory, for to every element of the one class there is an element in the other class that is its negation. Thus whenever logical content grows, informative content grows also and to the same degree.

Now the argument for the infinite size of the logical content of any theory can be presented as Popper presents it:

Let there be an infinite list of statements a, b, c, \dots which are pair-wise contradictory, and which individually do not entail t . Then the statement " t or a or both" is deducible from t , and therefore belongs to the logical content of t . From our assumptions regarding a, b, c, \dots , it can be shown that no pair of statements of the sequence " t or a or both", " t or b or both", \dots , entail one another. It then follows that the logical content of t must be infinite. (1976, pp. 26-27)

The following is a proof of the assumption that no pairs of the infinite sequence of statements entail one another.

The statement " b or t or both" follows from " a or t or both" if and only if it follows from a ; that is, if and only if it follows from " a and non- b ". But this last statement says the same as a (because b contradicts a). Thus " b or t or both" follows from " a or t or both" if and only if t follows from a ; and this, by assumption, it does not.

Popper argues that since the information content of any theory is infinite, we can never know all that we talk about. Thus since Einstein's theory is incompatible with Newton's theory it must be part of the information content of Newton's theory. Newton could hardly have been expected to know this. Indeed, there are an infinite number of complex, non-trivial theories which are part of the information content of Newton's theory.

" t " could be: "Communism will be realized when market monopolization has increased to the point where there is only one capitalist agency. And under present trends p , this will occur in exactly n years". (Where p is a specification of the characteristics of the trend, and n is a finite number.) Then the infinite sequence of pair-wise contradictory statements can be constructed by substituting $n+1, n+2, n+3, \dots, n+y$, and so on for the rest of the natural numbers. Clearly, however large n is there remain an infinite number of logically possible

years in which communism could be established. The same can be done with the hypothesis about trends, p. But all these theories, even though they are but a variation on t, constitute an infinite set and every one of them contradicts t.

Each of the above arguments is sufficient to show that Manning made a very serious error in saying that ideologists

...see all in the way of their belief and see all of what they believe. (1976, p. 141).

In putting his ideas into text the ideologist changes in a very special way the logic of his situation. He creates a set of ideas whose implications and ramifications go far beyond his comprehension and thus control. He cannot completely foresee how they will fare in argument, what sort of criticism they will provoke etc. In fact, this problem goes far beyond the capacity of any abstractly conceived predictor. In an important sense the ideologist becomes alienated from his own thought. (Bartley was well aware of this implication, but in some respects did not apply it to the evolution of an ideology under criticism. See especially "Alienation Alienated", chapter XVIII of his 1987). More importantly for our problem, no leader of a movement can control how the various propagandists will deal with criticism, specifically which immunising stratagem (in our example, which substitutes for n or p) will be created in response to each criticism.

The logic of the propagandist's situation is such that if an ideology is to survive and propagate, its chances of doing so are increased if its adherents actually propagandise. The various propagandists will then meet various counterarguments. But at least some of these counterarguments will be unpredictable. As Popper (1957) has argued, new ideas - which includes criticisms and defences - cannot be predicted. If the number of propagandists is quite small, then they may continually consult with each other about the appropriate response to each criticism. On the other hand, if the number of propagandists is significantly large, then their responses to counterargument cannot be controlled by some kind of democratic decision or from a centre. The responses must be improvised there and then in debate. The possibility is then open for different propagandists to improvise quite different "immunising stratagems". This will, of course set up the basis for the emergence of factions. But the gulf between such

various factions runs deeper. In understanding a criticism one is understanding the theory being criticised, either because the criticism brings out a previously unnoticed implication or because one sees that it does not. If different propagandists improvise different immunising stratagems, they are at the same time developing (perhaps overlapping but) different understandings of the original canonical theory. As is clear from the above example of a Marxist style t , a propagandistic theory allows infinite room for divergence of opinion on what figure, $n+y$, to substitute for a falsified prediction n . (The substitution may occur not in response to a failed prediction but as a necessary consequence of other changes in the overall doctrine.)

In addition, each new immunising stratagem brings its own problems. And since the argument is general, each problem may be solved in any of an infinite number of ways. So we have a rapid accumulation of possibilities for the emergence of factions, the various propagandists possibly holding increasingly divergent interpretations of the same symbolism.

12. *Conclusion*

A perfectly self-reproducing "mind-virus" together with its self-reproducing social system, is an extremely delicate and vulnerable system. The World-3 character of the mind-virus makes its reproduction through a population of minds profoundly unstable. This is so even when many corrective mechanisms are brought into play, principally because these repair mechanisms bring with them their own logical problems of reproduction. I believe that Campbell's emphasis on the blindness of the evolution of ideas is reinforced by my paper. But this reinforcement comes at a price, and the price is the quest to completely physicalize the three stage scheme of evolution plus the beliefs, expectations, and attitudes undergoing this evolution. If these beliefs have any complexity at all, they will exist partly as linguistically formulated doctrines - unfathomable World 3 entities. And these cannot be fully embodied neurologically or psychologically.

As a test case, I examined the most hopeful candidate for a self-reproducing belief: a system protected by a sequence of what Popper calls "immunizing stratagems". However, we found that a network of ideas

can evolve under the impact of refutation and criticism (acting as coselectors), with the elimination of error, without any explicit acknowledgement of error. Paradoxically, one of the ways this can happen is through the very use of so-called "immunizing stratagems". The class of Immunizing stratagems which have this structure are better seen as face savers, not theory savers. Immunizing stratagems do not prevent the critic from analysing the ideologist's doctrine, pointing to the immunizing stratagem involved as a criticism in itself, and then providing an empirical refutation of the reinterpreted doctrine. Therefore, immunizing stratagems do not serve to preserve an ideology for which they have been instituted.

Though there are barriers to sound argument, none are insurmountable. With enough ingenuity any computer virus can be eliminated. Similarly, all systems of ideas and their adherents are open to sound argument, albeit a long and difficult one. Nevertheless, we should expect error to be perpetuated even in a population of humans who infallibly eliminate error, because it may be passed on more quickly than it is repudiated.

The larger question of openness to argument and the role of rationality in the evolution of systems of ideas is explored in my forthcoming book *The Logic of Mind-Viruse*.

Lancaster University

NOTES

1. Popper (1934) had used the term "conventionalist stratagem". Cf. Hans Albert (1967). David Miller subsequently informed Popper of note 1 on page 560 of Arther Pap (1963). In this note Pap anticipates Popper's usage.
2. Mises continued his argument (1951, p. 135-163) and refuted more recent criticisms in his 1949 (pp. 194-711). With regard to "market socialism", see F.A.Hayek "Socialist Calculation III, The Competitive 'Solution', in *Individualism and the Economic Order*.

REFERENCES

- Albert H. (1967), *Marktsoziologie und Entscheidungslogik*. Neuweid and Berlin: Herman Luchterhand Verlag.
- Barone E. (1908), "IL Ministro Della Produzione Nello Stato Collectivista", in *Giornale Degli Econ*, II. Also published as "The Ministry of Production in the Collectivist State, in *Collectivist Economic Planning*, London, 1935.
- Bartley W.W. III, (1962), *The Retreat to Commitment*. 2nd English edition, Open Court, Chicago, 1984.
- Bartley W.W. III. (1987), *Evolutionary Epistemology, Rationality, and the Sociology of Knowledge*. eds. Gerard Radnitzky and W.W. Bartley III. 2nd printing, 1988.
- Boudon R. (1986), *The Analysis of Ideology*. Originally published as *L'ideologie: ou l'origine des idées reçues*, Paris, Librairie Arthème Fayard. Chicago: University of Chicago Press. Polity Press, Cambridge, 1989.
- Campbell D.T. (1979), "A Tribal Model of the Social System Vehicle Carrying Scientific Knowledge", *Knowledge: Creation, Diffusion, Utilization*, Vol. 1. No. 2, pp. 181-201.
- Campbell D.T (1987), "Neurological Embodiments of Belief and the Gaps in the Fit of Phenomena to Noumena", in *Naturalistic Epistemology*, Abner Shimony and Debra Nails (eds), pp. 165-192. D. Reidel Publishing Company.
- Campbell D.T. (1993), "Plausible Coselection of Belief by Referent: All the "Objectivity" That is Possible." in *Perspectives on Science*, 1, No. 1.
- Cioffi F. (1970), "Freud and the Idea of Pseudo-Science", in R Borger and F Cioffi, *Explanations in the Behavioural Sciences*, C.U.P. Reprint 1978.
- Dawkins R. (1976), *The Selfish Gene*, Penguin. Paperback reprint, 1990.
- Duhem P. (1914), *The Aim and Structure of Physical Theory*. Princeton University Press, 1991.
- Flew A. (1975), *Thinking about Thinking*, Fontana. Third impression, 1978.
- Gellner E. (1985), *The Psychoanalytic Movement*. Paladin Grafton Books. 1988 reprint.
- Lakatos I. (1970), "Falsification and the Methodology of Scientific Research Programmes", in *Criticism and the Growth of Knowledge*, C.U.P Reprint, 1979.
- Manning (1976), *Liberalism*. J M Dent & Sons Ltd, reprint, 1982.

- Marx K. (1867), *Capital*, Vol.I, First German edition. Lawrence and Wishart, London, 1974.
- von Mises L. (1935), "Economic Calculation in the Socialist Commonwealth". Reprinted in F.A.Hayek, ed. *Collectivist Economic Planning*. London. George Rutledge and Sons, pp. 87-130.
- von Mises L. (1949), *Human Action*, William Hodge & Company.
- von Mises L. (1951), *Socialism*. New Haven: Yale University Press.
- Musgrave A. (1978), "Falsification, Heuristics, and Anarchism". in *Progress and Rationality in Science*, Boston Studies in the Philosophy of Science, Volume LVIII, eds., G. Radnitzky and G. Andersson. Published by D. Reidel.
- O'Hear A. (1980), *Karl Popper*, Routledge & Kegan Paul. Reprint, 1982.
- Pap A. (1963) "Reduction Sentences and Dispositional Concepts", in Paul Arthur Schilpp (ed.), *The Philosophy of Rudolph Carnap*. La Salle, IL.: Open Court Publishing Co, pp. 559-97.
- Percival R.S. (1994), "Dawkins and Incurable Mind-Viruses: Memes, Rationality and Evolution", *Journal of Social and Evolutionary Systems*, Vol. 17, No. 3, pp. 243-286.
- Popper K. (1934), *Logic of Scientific Discovery*, Originally published as *Logik der Forschung*, Vienna. Hutchinson & Co Ltd., 9th impression, 1977.
- Popper K. (1957), *The Poverty of Historicism*, Routledge and Kegan Paul. Reprint, 1979.
- Popper K. (1958), "On the Status of Science and of Metaphysics", first published in *Ratio* 1, 1958, and is now chapter 8 of *Conjectures and Refutations*.
- Popper K. (1963), *Conjectures and Refutations*. Routledge and Kegan Paul. revised edition, 1969.
- Popper K. (1974), *The Philosophy of Karl Popper*, ed., P.A. Schilpp, Open Court, pages 1059-65.
- Popper K. (1974), "The Problem of Demarcation", Sections 5-8 of "Replies to my Critics", Schilpp volume.
- Popper K. (1976), *Unended Quest*, Fontana/Colins. Originally published as "Autobiography of Karl Popper" in Popper, [1974]. Revised edition, 1976.
- Popper K. (1982a), *The Open Universe*. Volume II of the Postscript to *The Logic of Scientific Discovery*. Hutchinson, editor W.W.Bartley, III.
- Popper K. (1982b), *Quantum Theory and the Schism in Physics*, Volume III of the Postscript to the Logic of Scientific Discovery.
- Popper K. & J. Eccles (1977), *The Self and its Brain*. Springer International.

Shils (1968), "The Concept and Function of Ideology". I.E.S.S. Volume VII.

Watkins J.W.N. (1984), *Science and Scepticism*, Hutchinson.

Watkins J.W.N. (1958), "Confirmable and Influential Metaphysics". *Mind* 68.

Wells G.A. (1988), *Religious Postures*. Open Court.