European University Institute

MAX WEBER PROGRAMME

EUI Working Papers

MWP 2010/37
MAX WEBER PROGRAMME

ON OBJECTIVE KNOWLEDGE IN SOCIAL SCIENCES AND HUMANITIES: KARL POPPER AND BEYOND

Chiara Valentini (Editor)

Ramon Marimon, Chiara Valentini, Simon Blackburn, Carol Cleland, Gerald Postema, Harry Collins, Justin Cruikshank, and Frédéric Vandenberghe (Contributors)

EUROPEAN UNIVERSITY INSTITUTE, FLORENCE
MAX WEBER PROGRAMME

Objective Knowledge in Social Sciences and Humanities: Karl Popper and Beyond

CHIARA VALENTINI (EDITOR)

EUI Working Paper **MWP** 2010/37

ISSN 1830-7728

Printed in Italy
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy
www.eui.eu
cadmus.eui.eu

Abstract

This collection of papers addresses the issue of ‘objective vs. subjective’ knowledge in the social sciences and humanities: how we may get ‘objective’ knowledge out of ‘subjective’ perceptions; how ‘induction’ and ‘deduction’ should interact; how we can make policies or legal recommendations based on ‘objective knowledge’; how social agents’ knowledge should be modelled.
Drawing on the structure of the conference, the papers are organized in sections which address a set of interrelated questions, against a common thematic background provided by Popper’s contribution on objective knowledge in the social sciences and humanities: the ‘induction problem’ and the accumulation of ‘subjective’ knowledge out of ‘objective’ knowledge; ‘objectivity’ of the law and of social policies; objectivity of facts and causal relations in the social sciences and humanities; the objective/subjective rationality of social agents.

Keywords

Social sciences, humanities, Karl Popper, knowledge, objectivity.

Table of Contents

Foreword

It is a pleasure for me to present this volume of selected papers ***On Objective Knowledge in Social Sciences and Humanities: Karl Popper and Beyond***. The papers are an outgrowth of most of the papers presented, the 13 March 2009, in ***The 3rd Max Weber Programme 'Classics Revisited' Conference***, which was organized by a group of Max Weber Fellows, and as the previous two *'Classics Revisited'* conferences, centered on a specific theme of interest in the Social Sciences and Humanities in the 21st Century, taking a leading 'classic' as a reference, but focusing more on the theme — in this case, the recurrent question of what is 'objective knowledge' in the social sciences — rather than in the 'classic' — in this case, Karl Popper; although this excellent collection of papers, edited by Chiara Valentini, is a contribution to both: the papers reassess the general theme of 'objective knowledge', as well as the specific contributions of Karl Popper.

Karl Emil Maximilian Weber (1864–1920) was, for an obvious reason, the first 'classic to be revisited'. As Fritz W. Scharpf reminded us in *The First Max Weber Lecture* (*The Inaugural Lecture of the Max Weber Programme*, the 4 October 2006) Max Weber set social sciences aside from natural sciences:

> Laws are important and valuable in the exact natural sciences, in the measure that those sciences are universally valid (…) For the knowledge of historical phenomena in their concreteness, the most general laws, because they are most devoid of content are also the least valuable (…) In the cultural sciences, the knowledge of the universal or the general is never valuable in itself[1]

Our second 'classic revisited' was David Hume (1711-1776) who, in his own reassessment of his famous 'induction problem,' said:

> In every judgment, we ought always to correct the first judgment, deriv'd from the nature of the object, by another judgment, deriv'd from the nature of the understanding[2]

Revisiting the twentieth century classic Karl Popper (1902-1994) provided continuity with our two previous conferences and it was a way to show how, following Hume's dictum, progress has been made since Max Weber divided cultural sciences from the natural sciences. As Justin Cruickshank says in this volume:

> The social sciences, like the natural sciences, do not get knowledge because they adhere to a fixed set of ontological definitions or because theories are able to map all the essential determinants of social reality. Instead, knowledge grows in the social sciences through substantive problem-solving.[3]

I do not want to dwell more on the topic and spoil the film with its trailer, but just to invite the reader to the works of the leading scholars collected in this volume. They are not futile attempts to resolve or close the problem of: what is objective knowledge in the social sciences?'. Instead, the collected papers are new contributions that enhance our understanding of the problem, a problem that in times of socio-economic crisis no social scientists can neglect.

---

[1] "Objectivity in Social Science and Social Policy" 1904/1949, in *The Methodology of the Social Sciences*. ed./trans. E. A. Shils and H. A. Finch. New York: Free Press.

[2] *A Treatise of Human Nature*, edited by David Fate Norton and Mary J. Norton, Oxford/New York: Oxford University Press, 2000

[3] "The Importance of Nominal Problems" (p. 71).

But before giving the floor to the authors, let me express to them my sincere gratitude, as well the other scholars who participated in the lively conference[4], and the Max Weber Fellows — Mathias Delori, Joshua Derman, Alexander Kriwoluzky, Miriam Ronzoni and Chiara Valentini — who organized the conference and, specially to Chiara for her excellent job in editing this volume.

Ramon Marimon

[4] Unfortunately, due to copyright issues, it has not been possible to include the contributions of Susan Haak (Department of Philosophy and School of Law, University of Miami), and David Schmilder (Department of Economics, the Ohio State University, and the School of Mathematical Sciences and the School of Business Administration, Tel Aviv Universiry).

Introduction

On Objective Knowledge in Social Sciences and Humanities: Karl Popper and Beyond

Chiara Valentini♣

On 13 March 2009, the annual Max Weber conference on the 'Classics Revisited' brought together prominent scholars and Max Weber Fellows to discuss 'Objective Knowledge in Social Sciences and Humanities: Karl Popper and Beyond'.

In line with the previous tributes to Max Weber and David Hume, the 2009 conference paid homage to Popper's thought and hosted an exchange of ideas on issues of crucial relevance for different fields of research and study: from economics to history, from law to sociology and political science, all with the inter-disciplinary perspective that distinguishes the Max Weber Programme research activities. From this point of view, the thinking of Sir Karl Popper (1902 – 1994) is particularly enlightening as it offers a powerful and comprehensive account of human knowledge, one that embraces the diverse domains of scientific creativity and lends itself to a cross-sectional reading. Popper, indeed, worked out the possibilities of human learning, and has provided us with a deep understanding of knowledge as well as of its use in solving practical problems. In this sense, Popper's epistemology and social philosophy are strongly interconnected: methodologically speaking they both require a critical and "fallibilist" approach in dealing with the "real world"; as for the contents, both science and society are depicted by Popper as "open" to criticism and to a pluralism of ideas and perspectives.

Popper's social philosophy developed in parallel with his vision of science under the sign of a unity founded on critical reason, in a comprehensive philosophical system that has fascinated and divided the intellectual community. In spite of weaknesses and disputed conclusions, Popper's critical rationalism was a turning point in the enquiry into the nature of human knowledge. It gave rise to a new perspective, with a strong inter-disciplinary appeal, due to the universal relevance of the ideas it fosters; especially the idea of critical reason as the engine and instrument of human learning and the driving force of research.

In the first part of this introduction, I sketch out a few of the conceptual and methodological contributions of Popper's "critical rationalism" to the social sciences and humanities. In the second part, I sum up the contributions to the conference collected in this volume.

Fallibilism and Critical Rationalism in the Social Sciences and Humanities

Popper's account of science combines the rejection of the verification principle with the use of falsification as the criterion of demarcation between science and non-science, and an emphasis on fallibilism with the quest for critical scrutiny and "openess" of thought in any process of discovery.

In all these respects, Popper's epistemology challenged the model of scientific learning that had been dominant in the history of Western science for many centuries, "certainly well into the nineteenth"[5]. According to this traditional model, which Roger Oldroyd has rendered by the metaphor of the "arch of knowledge", scientific knowledge is the product of cognitive processes, which move upwards, from facts to hypotheses, and downwards, from hypotheses to facts, so as to design an "arch" whose structure is made of inductive and deductive segments and presents a certain "strength and security"[6].

For Popper, by contrast, scientific learning does not entail inductive passages, but is essentially deductive and proceeds from the formulation of conjectures towards their corroboration,

♣ University of Bologna. Max Weber Fellow, 2008-2009.

[5] R. Oldroyd, *The Arch of Knowledge: An Introductory Study of the History of the Philosophy and Methodology of Science*, London : Methuen, 1986, p. 363.

[6] R. Oldroyd, *The Arch of Knowledge: An Introductory Study of the History of the Philosophy and Methodology of Science* , cit.

through testing procedures aimed at selecting the ideas resistant to critical scrutiny and clearing the way for ideas which do not lend themselves to criticism.

Although this hypothetico-deductive turn has been harshly questioned, it undoubtedly offers a striking vision of the growth of knowledge. Crucial to this vision is Popper's shift from the verification principle to that of "falsification", as a criterion of demarcation between science and pseudo-science. According to this criterion, the ideas advanced in a scientific discourse acquire the dignity of scientific theories if they can be falsified, that is, if they have a sufficiently precise and explicative content that can be questioned and challenged. Scientific methodology, accordingly, does not consist of verification procedures, but of falsification procedures by which scientific conjectures are subjected to attempts at refutation.

Falsifiability, thus, is the key to Popper's answer to the problem of the nature of scientific knowledge, meant as a process of human learning as well as the product of this process.

In this perspective, our attitude towards cognitive claims plays a crucial role in their assessment: in line with falsificationism and conjecturalism, we should avoid justificationist approaches and take a fallibilist position.

What scientific learning needs, in fact, is the awareness that "though we may seek for truth, and though we may even find truth (as I believe we do in very many cases), we can never be quite certain that we have found it. There is always a possibility of error".

Fallibilism, in this sense, is the link between Popper's philosophy of science and his vision of history and society; it is the core of a comprehensive theory of knowledge that renders any form of human learning an open-ended and critical process of advancing and testing hypothetical solutions to problems.

The inter-disciplinary appeal of this theory lies in this declination of critical rationalism in terms of fallibilism, which, in Popper's strategy, stands as the substitute for certainty against any problem faced by rational agents.

All the learning processes oriented towards discovery, in fact, require a "scientific attitude", that is, a willingness to advance ideas and expose them to criticism and revision. This experimental approach applies beyond the domain of natural and empirical sciences, and extends to social life and history, where the sense of fallibilism turns into the awareness of the unpredictability of the course of events and of the open solution of the problems faced by social agents. Critical rationalism, here, requires the rejection of holistic approaches and the adoption of methodological individualism as part of the "critical" assessment of social, political and historical issues.

"Critical" research, in all the forms it takes, must be rigorous in bringing forward ideas as much as rigorous in evaluating arguments and counter-arguments, in search of possible errors and weaknesses. A reliable researcher, indeed, wants to "learn from mistakes" and reliable knowledge is the product of this never-ending learning.

For all these aspects, Popper's philosophy transmits to the social sciences the idea that human knowledge is not a "kingdom" of truth, but rather a problem-solving effort that can generate reliable theories and "objective" ideas.

In a fallibilist perspective the conjectural and uncertain process of human learning cannot reach "perfect knowledge" and give definitive solutions to scientific problems or social questions. Rather, it can reach an uncertain, provisional, imperfect knowledge about the real world by yielding ideas that can compete with other ideas and resist critical scrutiny.

In the absence of truth, a critical attitude allows us to arrive at "objective" solutions, which are sufficiently explicative and corroborated according to the method of trial and error.

From this point of view, objectivity is a question of method, the result of "critical" courses of research. These must be oriented towards the enhancement of knowledge and evolve from the dimension of subjective experience to the dimension of discovery, in which our ideas release themselves as a product of the human mind and become part of "world three". The conceptual category of "world three" encompasses the independent world of "problems, theories, critical arguments", which have an autonomous existence, and is meant to return the objective dimension of

shared knowledge to us.

As outlined by Neil MacCormick[7], this vision of human learning as a "testing" process of discovery oriented towards reliable and objective solutions, draws our attention to the importance of the "whole" body of shared knowledge, which allows us to "see" problems, become aware of them, reason upon tentative solutions, and assess these solutions against others.

In this sense, Popperian experimental logic requires testing procedures which do not scrutinize hypotheses "in vacuo", but involve the "reliance on auxiliary hypotheses" embedded in "what is already known". The advancement of knowledge, thus, benefits from "discoveries" "involving extrapolation" from existing knowledge, with which "they are compatible" and "with which they make sense". In this perspective, research, from that regarding natural phenomena to that concerning social questions, is "mobile", that is, not producing fixed systems of theories, but a continuum of ideas on ideas, by looking behind and beyond, in the understanding that no solution to a question is definitive. Furthermore, research is "open", that is, not producing isolated and self-reliant ideas, but ideas embedded in a context of knowledge, which supports our conjectures and their corroboration or rebuttal.

The legacy that Popper has left to the social sciences and the humanities lies not only in the critical attitude towards research, but also in the account of human beings as rational and open-minded agents. The "experimental" approach to research, in fact, is grounded in a conception of individuals as free and rational agents retaining a critical "power".

This conception underlies Popper's vision of social studies and social life, being essential to a rational method of social research as much as to the "open" society of free individuals that he defends. Both of them, indeed, demand the diversity of ideas and perspectives, the exchange of reasons, the accurate scrutiny of the diverse solutions advanced in all the discourses aimed at questioning scientific or social or historical issues.

In this respect, Popper's epistemology turns into an entire philosophy of knowledge that is applicable to any dimension in which human reasoning is engaged: epistemic fallibilism is a requisite of both the "open knowledge" and the "open society", which simply cannot be rational or democratic without the awareness that there are no "true" or definitive solutions to the problems faced by social agents.

The "open society", according to Popper's methodological individualism, is highly individualistic and ruled by critical thinking; it is a society in which human beings are called to face problems, reason upon them, envision and compare possible solutions, choose and be responsible for their choice.

Each of these steps must be "critical"; at each of these stages individuals are entrusted with the task of searching for their own errors and questioning a *status quo* that is always "open" to criticism and change.

From this point of view, Popper's contribution to the advancement of the social sciences and humanities is more than methodological as his vision of society and history promotes a specific account of social life. The development of individuals' critical power, indeed, needs a liberal, democratic, individualistic framework.

Critical rationality, here, works for the "falsification", via inter-subjective "verification", of arguments and counter-arguments, and serves as the standard of procedural correctness in open and democratic decisional processes. Popper's trial and error perspective, thus, applies to society, whose assets are taken as always revisable in a never-ending adjustment within a transparent public discourse among rational, accountable, critical agents.

In this sense, we may find some points of contact between the Popperian idea of an open society and the "discursive" perspective embraced, in contemporary debate, by theoretical models propounding a deliberative account of democracy. These models are constructed around the idea that decision-making processes in the public sphere should develop in an open, transparent and "critical" way. Both Popper's critical rationalism and the quest for a deliberative exercise of public reason

[7] N. MacCormick, *Legal Reasoning and Legal Theory*, Oxford : Clarendon Press, 1978, pp. 101-102.

assign a crucial role to the exchange of ideas, which presupposes the opportunity for the individuals involved to defend their own positions and also to take into account and welcome any objections. Popper's philosophy, in the end, gives to human knowledge the possibility of correcting and changing things and shows the *way* for this change: confronting problems with a fallibilist approach, in an endless search for possible solutions and counter-solutions within a social and political context that, like science, must be a theatre for the transmission and exchange of reasoned proposals, ideas and arguments.

From this point of view, Popper's contribution to social sciences and humanities has been highly remarkable for it brings to light the creative force of research and the capacities of human knowledge, reason and life.

This collection of papers outlines and addresses the significance of Popper's philosophy for the social sciences and humanities in relation to several, and crucial, questions: how we may get 'objective' knowledge out of 'subjective' perceptions; how 'induction' and 'deduction' should interact; how we can make policies or legal recommendations based on 'objective knowledge'; how social agents' knowledge should be modelled.

Drawing on the structure of the conference, the papers are organized in sections which address a set of interrelated questions, against a common thematic background provided by the issue of Popper's contribution on objective knowledge in the social sciences and humanties: the 'induction problem' and the accumulation of 'subjective' knowledge out of 'objective knowledge'; 'objectivity" of the law and of social policies; objectivity of facts and causal relations in the social sciences and humanities; the objective/subjective rationality of social agents.

The first section, titled 'From 'Subjective' To 'Objective Knowledge': The 'Induction Problem' Revisited' includes the contributions of Simon Blackburn (Department of Philosophy, University of Cambridge) and Carol Cleland (Department of Philosophy, University of Colorado).

Blackburn's paper, 'Popper and His Successors', outlines two views of Popper's philosophy of science. According to one view, far from solving Hume's problem, Popper fails even to recognize its starting point, hence we must return to Hume. According to the second view, by retaining at least an objective conception of falsification, Popper himself is overtaken by Kuhn, and post Kuhnean radical philosophers of science such as Feyerabend or Hacking. In this perspective, Blackburn argues that while Popper may need more Hume and more Kuhn in his eventual philosophy, there is potential for us to build on his insights and look for a synthesis.

In her contribution, 'Common Cause Explanation and the Asymmetry of Overdetermination', Carol Cleland argues that the methods of prototypical historical science differ from those of classical experimental science, and that these differences in practice are underwritten by a time asymmetry of causation (David Lewis's "asymmetry of overdetermination") that provides the needed justification for the principle of the common cause. According to the first half of the asymmetry of overdetermination, the present is filled with epistemically overdetermining traces of past events; hence it is likely (but not certain) that a puzzling association (correlation and/or similarity) among present-day phenomena is due to a last common cause.

In the second section, titled 'Objectivity of The Law and of Social Policies', the contribution of Gerald Postema (Department of Philosophy, University of North Carolina, Chapel Hill) on 'Hayek and Popper on the Evolution of Rules and Mind' argues that Popper's view of the evolution of the realm of objective knowledge and Hayek's notion of spontaneous order offer insights into the nature and emergence of social norms and focuses on these insights and their implications for our understanding of the foundations of law.

The third section, titled 'Objectivity of Facts and Causal Relations in the Social Sciences and Humanities', includes the contributions of Harry Collins (School of Social Sciences, Cardiff University) and Justin Cruickshank (Department of Sociology, University of Birmingham).

Harry Collins addresses the issue of 'Demarcation Criteria and Elective Modernism' and

argues that the Popper's `falsifiability' criterion of demarcation, along with all the attempts to solve the demarcation problem for science, have failed because the problem itself was misconceived: the problem is sociological rather than logical; it is a matter of describing the `form-of-life' that constitutes the `family' of sciences. Conceived this way, Collins argues that all attempts at demarcation, including Popper's, have actually been successful. In this perspective, the discomfort we feel — we know the demarcation criteria make sense even though they can be shown to be flawed — is thus resolved.

The contribution of Justin Cruickshank on 'The Importance of Nominal Problems', focuses on the criticism of general theory advanced by neo-pragmatists, some post-Wittgensteinians and some feminists. Much of this criticism holds that general theory entails the construction of a closed system of abstractions that are divorced from the practices of agents and which seek epistemic justification by capturing the essential properties behind the fluid actions of agents. Cruickshank argues that whilst neo-pragmatists have correctly identified some problems with general theory, their solution is incorrect. One response to their criticism is to reject 'theory' in favour of a focus on an agent's creativity. Drawing on the work of Popper on methodological nominalism, problem-solving and the replacement of justification with criticism in epistemology, Cruickshank develops this argument and claims that rather than juxtapose theory to creativity, theory may animate social scientific creativity.

The fourth section, devoted to the issue of 'Modeling Individual and Social Agents as Objective/Subjective 'Rational' Agents', includes the contribution of Frédéric Vandenberghe (Instituto Universitario de Pesquisas do Rio de Janeiro) titled "Falsification Falsified. A Swansong for Lord Popper".

Vandenberghe claims that Popper's neo-positivism damaged natural sciences: as Popper had to admit that the 'covering law-model' does not really apply to the social sciences, he developed an alternative model of explanation for the human sciences and introduced the situational logics of rational choice as second best in the human sciences. Vandenberghe, thus, submits Popper's critical rationalism to a metatheoretical critique and questions its ontological, epistemological, ideological, ethical and anthropological presuppositions from a realist, phenomenological, hermeneutic, communicative, humanist perspective.

Popper and his Successors

Simon Blackburn[♣]

I

I wish to start by reflecting a little on the problem of induction, and Popper's response to it. As is well known, Popper believed he could dismiss the problem, and offer a philosophy of science that bypassed it. I do not believe that is possible, and I think Popper's attempt to sidestep it opened the door to ever more debilitating philosophies of science in the second half of the twentieth century. I also believe that with the general retreat of postmodernism and scepticism, we face the need for a new approach, if one can be found.

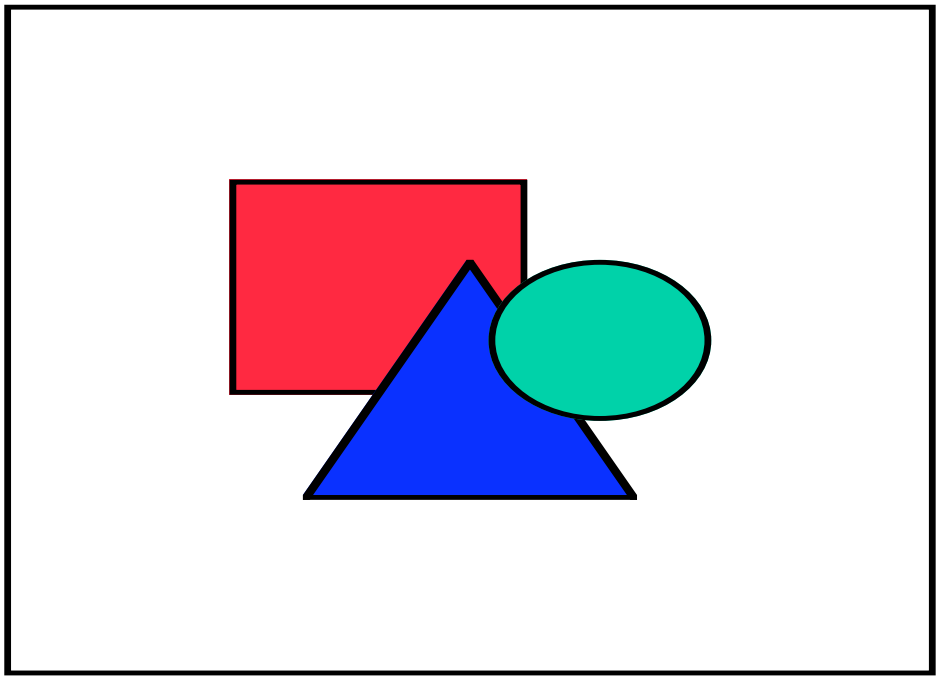
As the Roman writer Cicero reflected, times change and we change in them. The world goes round, and things alter. But not too much. There has to be a speed limit, and fortunately there is. Our domestic cat will not suddenly start talking. It won't even change into a dog overnight, and neither will I walk through a wall, or grow another pair of arms. In fact we think change only goes on in the comfortable shelter of things which do not change: laws of nature, that determine the ongoing pattern of things.

So what can reason tell us about these uniformities in nature? Here we meet an impasse. It seems we would have to give a reason for any constancy that we find, either by relying on empirical evidence, or by relying on something like mathematics and logic: either 'a posteriori', after experience, or 'a priori', in advance of experience. However it seems as though all that our a posteriori knowledge could tell us is that at least if some things (gravity, strong or weak forces, some very finely tuned laws of nature) keep on in their old familiar way, then other dependent things will do so. If gravity continues in the way we know, then the solar system will continue its revolutions. If the strong and weak forces within the nucleus of atoms continue to work as they have done, matter will not fly apart, nor implode. But we can only argue for any one uniformity by relying on another one, until we come down to fundamental magnitudes determining such things as the strength of charge on the electron, the speed of light, or the strengths of electromagnetic forces. These seem to stay put, but why should they do so? Any a posteriori reason will simply push the problem back onto some other pattern, onto which we then clutch. If we ask, in turn, why that keeps on keeping on, eventually we come to the point where there is no answer. It just seems to do so. It has done so wherever and whenever we have investigated, perhaps, so we extrapolate. We are confident that it is reliable, and will continue to be so. But then the question is whether this is any more than an article of faith, an unsupported dogma on which all our scientific constructions ultimately rest?

Perhaps pure reason can help. But as Popper well knew there is no a priori reason we can assign why radical change, chaos even, should not break out. What we would really like is a necessity, a straightjacket on events which (logically) *cannot* change. It would have to be something timeproof and self-sustaining, a law which once written down, cannot be revoked. The ancient image of Atlas sustaining the world on his shoulders might give us a mythological glimmer or allegory of what we would like, but of course anything analagous to human fortitude is not going to be immune to time and change. Atlas, for all we can understand, might get bored or tired or distracted. He might shrug, and drop the whole shooting match. So to get what we want, we need a fact of a different kind altogether, and then the fear arises that we have no idea what such a thing could be: our understandings cannot comprehend it.

Unfortunately the same logical difficulties beset Popper's own preferred alternatives of conjecture coupled with falsifiability. Thus imagine a graphical demonstration of the problem of induction. Here we are plotting the value of some potential variable through time, and hope to make a prediction as to its future value or values across some interval of the future:

[♣] Simon Blackburn is Professor of Philosophy at the University of Cambridge



We want a reason for confidence in the straight hypothesis, as opposed to any of the continuum many rival routes through the same space of possibilities. But each of these hypotheses is equally bold (each provides one value for any temporal point) and each is equally falsifiable. A priori we can see no reason for confidence in one rather than another. This is, incidentally, quite independent of Goodman's paradox, which adds the further reason for scepticism that if we redimension the graph, the hypothesis which in this scale looks 'straight' may look bent, while some hypothesis which in this scale looks bent may look straight. Goodman only tells us that if 'straightness' then ceases to be a logical property but appears to be more a property of our choice of language or our choice of dimensions, the hope of a logical reason for preferring S vanishes altogether. But even without accepting Goodman's belief that it is only convention and linguistic history that lead us to prefer one dimension to another, we can see that boldness and falsifiability tell us nothing that distinguishes all the rivals.

It is wrong to exaggerate this point, and this is where (together with David Hume) I begin to part company with Popper. It is not an encouragement actually to worry about our immediate futures. Our lives are premised on the supposition that the immediate future will indeed resemble the immediate past. Our best guide to what to eat is what we have successfully eaten in the past; our best guide to the number of limbs we will have when we wake up, or the language we will speak or the place we will be, is how we were when we went to bed. We make structures of steel not iron because steel has always shown greater strength under tension; we expect to require oxygen and water in the immediate future just as we have always done. Anyone thinking these regularities are about to break in his favour (or to his harm, more likely) is deluded. Popper was famous for asserting that all that science could give us are "bold conjectures" as to what might happen. But if the right attitude to a bold conjecture falls short of actually believing it, the comparison must be wrong. Our empirical science, our discoveries about the way the world works, give us more than mere hypotheses or mere conjectures. They give us our certainties, the beliefs which our whole lives presuppose. Across the whole lanscape of our lives, the straight hypotheses demand our confidence, while the bent ones do not.

In fact, as Hume saw, the philosophical sceptic arguing that we should not place any confidence in these continuities is wasting his breath. Nature forces us to expect things as we do. I cannot jump off a cliff without expecting to fall, or deliberately walk into a wall without expecting to be stopped, any more than a dog or a cat can. Our animal natures tell us how to navigate our world.

They make us confident, and no reasoning could ever undermine it. Unless, perhaps, some scientist got wind of a cataclysmic change on the way, by himself relying on yet more uniformities whose relentless grip is about to destroy that of gravity, or the adhesion of matter, or the other forces which keep our lives in order. And in that case, perhaps, we might not know what to think.

Now, not only are our lives premised on uniformity, but I would argue, our very capacity to think at all is so founded. I believe, with Kant, Wittgenstein, Sellars, and many others that this capacity requires us to conceive of ourselves as occupying a point of view on a spatially extended, external world. But that in turn requires believing, for instance, that things continue in the same way independently of my experience. I am not at present in Oxford or Moscow, or for that matter in the room next door to where I write. But my conception of myself as in a public world requires that I believe that Oxford, Moscow, and the room next door exist, all the same. Now my only reason for believing such a thing is that uniformity requires it: an expectation of catastrophic change in Oxford or Moscow or the room next door would negate the expectation. And finally my only two sources of evidence for that uniformity is what the world has exhibited so far, and what it is exhibiting here where I am. If extrapolation of those were 'bold and conjetural' then my conception of myself as inhabiting a space would be similarly bold and conjectural.

In some contexts the inevitable confidence in regularity can falter, and here Popper's description of us as merely making bold conjectures can do real damage. Consider that the standard timeline for cosmology and geology, the age of the earth, the formation of rocks, and the evolution of animals, is premised on regularities. The regularities include those of a variety of kinds of radioactive decay, extrapolations from rates of deposition and rates of formation of rocks and continents, and other scientific techniques. These can be coordinated and calibrated, and we can use them to determine that the earth is some four billion years old, and then give a time scale for the events in the geological record. But if we say that all of this is, nevertheless, merely bold conjecture we open the way for biblical fundamentalists and creationists to say that their "bold conjecture" that the earth is in fact only six thousand years old is just as good a "hypothesis" as that which science gives us. What we need to say instead is that the creationist, just as much as the scientist, premises his life and his activities on regularities—only he then maintains the right to "pick and mix" which ones he will choose and which he will not. His position is no better than that of someone saying that the world began five minutes ago, or that the creationist's holy book was written last week by extraterrestials from passing flying saucers, or that by flapping his hands he expects to fly. Once reason goes to sleep, there is no telling where we may end up.

It is, of course, usually true that the creationist and others like him faces no possibility of falsification. There is nothing that would lead him or her to face the necessity of revising their chronology. Whereas the scientist is by comparison vulnerable: a discovery that the radiochronometric measurements give radically different results, for example, would unseat their calibration and force a rethink in one direction or another. It is, however, unclear how much of an advantage this delivers, in terms of reason or rationality. A creationist might in principle admit that if, say, different books of the Bible led to conflicting ages for the earth, he too would be forced to rethink. He does not expect this to happen, but then neither does the scientist expect his measurements to break apart in the way I sketched.

Nevertheless, it is unnerving if these confidences are fundamentally groundless, relying on an unargued and unsupportable faith in a uniformity of nature that, for all we can see, might snuff out at any random moment. It would be much more comfortable if we could conceive of a straightjacket, something that is itself immune to the very possibility of change, and that in turn constrains nature to roll on as it always has done. In other words, we would like the laws of physics and chemistry and biology to be something like the laws of mathematics. Just as the law that between every pair of consecutive even numbers there lies an odd number is immutable, immune to time, necessarily true not just in the world as it happens to be but in any other possible world we can imagine, so we would like to find a constraining fact, a physical or metaphysical directive, ensuring the continuing good behaviour (from our point of view) of the natural order.

Unfortunately the best candidates physics can find for such a guarantor are more things that just keep on keeping on. These include the constant strengths of fundamental forces and magnitudes in

nature. In his book of that title astronophysicist Martin Rees describes 'just six numbers' on which the course of nature as we know it depends. They include the ratio of the electrical forces that hold atoms together to the force of gravity (about 10^{36} to 1), the number defining the amount of energy released when hydrogen fuses to create helium (.007 of its mass) and other magnitudes, each of which has to be just as it is, to within minute tolerances, if the orderly cosmos is to exist. Yet so far as we can see, such constants could in principle have been different, and could in principle change. Indeed, tests and measurements have been conducted on whether they *have* changed. For instance, it was speculated by distinguished physicists that the so-called 'fine-structure constant' which determines the strength of interactions between charged particles and electromagnetic fields has in fact changed its value a little over time (at present it stands at 1/137.03599958). Perhaps fortunately, in 2004 it was announced that so far as astrophysicists could tell, it has not. But there was never any suggestion that it simply *could* not have done so, like the structure of the numbers. Yet it would not take sophisticated astrophysical observations to determine that between any two consecutive even numbers there lies an odd number.

It is notable that if some measurement decided that such a constant had changed, the search would be on for something explaining the change. How would that proceed? It would have to find some other constancy which did not change. That is the way explanation works. So, for instance, it might be that the fine-structure constant has a value which depends in some law-like way on something else, such as the amount of energy in the universe; then that in turn becomes a fixed point, an unchanging law, and the same old question arises: what on earth or in heaven assures that *this* relationship doesn't change? Faced with this treadmill, David Hume remarked that the utmost that natural science could do is to "stave off our ignorance a little longer".

Some distinguished scientists argue that the fine-tuning that these fundamental constants show is so vanishingly improbable, such an extraordinary set of coincidences underlying the good behaviour (so far) of our world, that we must look to a divine explanation, both of the fortunate magnitudes they take and their apparent stability. If the timeproof straightjacket cannot be found *within* nature, then perhaps it is best thought of as lying *outside* nature. This is a new version of very old arguments for the existence of a benevolent deity, guiding and sustaining the good behaviour of nature. A new Atlas in fact: a deity who is not only the first cause and architect of the whole show, but also its sustaining cause or ground, without whose firm control the whole cosmos might spiral into a void of timelessness and chaos. But that way lies a blank: we have no conception of any candidate for suh a fact.

We might derive some consolation from the thought that if our confidence is ever betrayed, then at least we will have no knowledge of it. Our existence is entirely dependent on the delicate adjustments that keep on keeping on. If they fail, then in a twinkling everything is over. Perhaps if we can accept the notion of time itself requiring the ticking clocks of the cosmic order, then if that fails, time itself comes to an end with it. In that case we could have the consolation that natural regularities lasted forever, that is, there was no time at which they did not hold. But it is rather cold comfort. We would hope that if the constancies last for ever, then at least they will last beyond, say, next Wednesday. Being told that they last for ever, until the end of time, but that unfortunately next Wednesday will never arrive because time will cease to exist on Tuesday, is scarcely the same thing.

II

I now want to turn to the state of philosophy of science in the years following Popper's writings. But I am not going to take us through the various contributions of Kuhn, or Feyerabend, or Lakatos, but to the atmosphere which, I am sorry to say, they contributed, willingly or not. I shall take the notorious Sokal hoax, the most celebrated academic escapade of our time, as my touchstone. Everyone is also likely to know the outline: how in 1996 the radical "postmodernist" journal *Social Text* published an article submitted by Alan Sokal, a mathematical physicist at New York University, with the mouthwatering title "Transgressing the boundaries: towards a transformative hermeneutics of quantum gravity". As we know, Sokal then revealed the article to be a spoof, a tissue of nonsense that he had painstakingly assembled in order to parody the portentous rubbish that flew under the colours of postmodernism. By publishing it the emperors of that tendency revealed themselves to be as naked as the rest of academia had always suspected, and with this one coup Sokal himself became the toast of the town, a celebrity and hero of the resistance.

Before 9/11, the story goes, academe allowed its "anything goes" tendency to grow unchecked. With long prosperity, the disappearance of the Cold War, and of any great causes to substitute for it, a certain playfulness, an ironic, aesthetic and disengaged attitude to life and history was quite tolerable. This was leisure time, and we did not need too much self-scrutiny, nor nervous and serious books about who we are and what we stand for and where we may be heading. The relativist could hold court as the lord of misrule. You disagree with me? Whatever. That's your view, and who's to say? I expect it is true for you. It didn't do to thump the table or insist too much: philosophers, it was supposed, had taught us to see any such exhibition as nothing more than a bid for power, a rhetorical trick for imposing on others, and as such rather bad manners. Especially it would not do if those thumped at were victims of the colonial past, or descendants anxious to claim the status of victim. In that sector respect was the order of the day, even if it meant smiling politely at Creationist timetables of earth history, Hindu versions of science, homeopathic medicine, and any other stumbling pre-scientific attempt at understanding the world. In fact the only proper targets of disrespect were those "metaphysical prigs", as Richard Rorty described them, who wanted to keep the inverted commas off words like truth, reason, or knowledge.

The present decade is different. We have learned that disagreement matters, and that if our grasp of what we need to defend is feeble enough, there are people out there only too happy to wrest it away from us. We have learned that there is not much common reason that is everyone's birthright, and that when disagreement comes people cannot afford to shrug. There are times when we have to do better than "whatever" and "anything goes". A country needs to understand what is good, and also what is not, about its preferred ways of living; it needs to understand what is good, and why, about its science, history, and self understandings, and it even needs to understand what was good, and why, about the politics and ethics and ideals it has, let us hope temporarily, abandoned. When academe finds a postmodernist White House where the President and his advisers sneer at the reality-based community, then carnival time is well and truly over.

I greatly enjoyed the hoax. There is nothing in academic life more sickening than people pretending to understand things that they do not. Who cannot want to explode the long lines of intellectuals posing as having a close acquaintance with iconic items of twentieth-century progress—relativity theory, of course, but also quantum mechanics, set theory, Godel's theorems, Tarski's work on formal logic, and much else? Ridicule is exactly what is needed.

Nevertheless, I found myself not quite as wholehearted as some of my colleagues. The editors, like some feminist critics of the institutions of science, thought that the complacency of science impeded radical progress. They also believed, probably rather vaguely, like most of us, that twentieth-century developments in science showed examples that shook off deeply entrenched complacencies or prejudices. They had heard of Einstein, of course, and knew something of the destruction of the classical notion of simultaneity, or of the discovery of relativity of motion to an observer; they might have heard of Max Born or Niels Bohr and Werner Heisenberg, and the idea that at the fundamental quantum level, what is observed is a function of whether it is observed, so that there is no legitimate notion of how things stand independently of whether they are observed. This has been contested, certainly, and Einstein himself ran a long and ultimately rather futile campaign against it, just as he did against the indeterministic implications of quantum theory. But both are still centre-stage in the philosophy of physics. Finally, and most importantly, the editors, and postmodernist writers in general were well aware of the dominant authority of science in every part of our culture. What humanists say doesn't much matter, but what the men in the white coats say goes. They were probably, like the rest of us, somewhat ambivalent about that, on the one hand mistrusting the absolute sway of science, but on the other hand eager for some of its gloss.

Against this background in came a paper by an accredited mathematical physicist teaching at a very highly regarded university. And the whole tendency of the article was to confirm the view that developments in science, right up to the contemporary scene, could indeed hold messages that were useful for their radical hopes. The science was presented as confirming for them that things lie, more than we might suppose, in the eye of the beholder. And that was important to the aims of the journal. True, the editors cannot have understood a lot of the alleged physics, partly because there was actually nothing there to understand. Still, if a Professor at NYU couldn't get that stuff right, who could?

I do not find it so surprising that they ran with it. Should they have got it refereed by mathematicians, physicists, and set theorists? I am not sure. It is better to do so, no doubt, and I expect that the poor editors have woken up every morning since wishing that they had. But there are costs of time and effort in finding referees, and as often as not you end up with two things to judge rather than just one. Anyway it was the purported message of the physics, not the details, that mattered to their interest in it. And you do trust academics to get their own subjects right. For example, when I edited *Mind* if a paper had come in from a well-regarded historian in an eminent department, showing, for instance, that various facts about Hobbes's political experiences in Venice explain his attachment to some doctrine in political philosophy, I would have had to estimate the political philosophy myself. But I might well have taken Hobbes's presence in Venice as given: surely any half-way decent historian wouldn't have developed the point if he hadn't got that bit right? Almost certainly I would not have got the history refereed, even if I had known who to approach.

I also found something a shade distasteful about the position of those triumphalists who were crowing about the hoax. Very few of them would be able to make head nor tail of a page of any contemporary physics journal. So when Sokal tells them that some sentences in his hoax were physically perfectly correct, while others were egregiously false or nonsensical, they have to take it on trust, and this alone puts them in a rather poor position from which to crow over the hapless others who took all of them, including the wrong ones, on trust.

Still, if you do not know how to tell a counterfeit coin from a true one, you should not go around pretending that you do. This was not the worst vice of postmodernists even if it was one of them. Far worse vices included the penchant for unintelligible writing, for drawing wild inferences, and for throwing around irresponsible claims, such as the wonderfully absurd assertions that before tuberculosis was identified you could not die of it, or that Newton's laws of motion make up a rape manual. These are also in Sokal's sights, and compared with them the pathetic conceit of decorating writings with a pretence of acquaintance with mathematics and physics is relatively minor.

It is natural to say that postmodernist writings displayed a contempt for truth, and Sokal skilfully defends a fairly modest kind of scientific realism, reaffirming the claims of science to give us the truth, or at least to put us on its track. And he is justly scornful of postmodernist philosophers who appeared to denigrate truth: Richard's Rorty's old campaign to substitute "solidarity" for truth is, rightly, a particular target. Gaining the agreement of our fellows is not the same as getting things right. It only begins to approximate to it if our fellows are trained in observation, evidence, and theory, and even then agreement stands ready to be struck down by the arrival of yet further observation, evidence, and theory in turn. Sokal is very good at this, and although he disclaims any special philosophical expertise, he writes well about the philosophy of science. He is good at articulating a basic epistemology for science, against the scepticism of Karl Popper as much as the wilder constructivist writings that followed him.

On the other hand Sokal is certainly not the kind of warrior in the "science wars" who disdains each and every attempt to say something interesting about the historical, sociological, and cultural matrix within which science has taken place. My own view is that such warriors do a terrible disservice to science, and in particular to scientific education. I like to illustrate this with an event in my own daughter's education. She came back furious one day from her very good and very expensive school, announcing she was fed up with science. I asked why. Apparently the class had been told to solve some equations governing the motion of the pendulum. In particular they had been told to use the equation of potential energy at the top of the swing with kinetic energy at the bottom, to calculate the velocity at the bottom of the swing. I asked what the problem was. She said she didn't see what this so-called energy was. I asked if she had raised this with the teacher. She said she had, and had been told to get on and solve the equations. She never pursued any science again.

Yet if you look at the history of the pendulum from Galileo's work at the end of the sixteenth century, you find a wonderful story of ingenuity, of mathematics, of contested observations, of problems of trade and the need to find the longitude, of the gradual evolution of the calculus, of debates about whether "force" should be thought of as proportional to velocity or square velocity (which set Newton and Leibniz at each other's throats). A century later, there were yet more disputes involving Carnot, Joule, and Helmholtz about the relationship between work, heat, and energy. You do

not find the conservation law in the form of the equation that was tossed at my daughter until the eighteen sixties. And as an aside, it is a pretty silly place to start in explaining anything about the pendulum, since energy depends on mass, and Galileo asserted, right at the beginning, that the period and velocity of the pendulum are independent of its mass.

Such dogmatic, stupid teaching not only loses bright children to science. It means that the ones who remain have been spoon-fed a bunch of results and techniques with no understanding of how they were hammered out, nor what their birth-pangs were. This disqualifies students from understanding the epistemology of science, and therefore of engaging effectively with doubters and deniers, whether the issue is one of the age of the earth or the measurement of its temperature They may suppose that science speaks with one voice, and the only dissenters must be luddites like the notorious Cardinal Bellamine, who allegedly refused to look through Galileo's telescope, whereas the truth is that many of Galileo's assertions, including those about the pendulum, were contested by careful observers, including amongst others Descartes and Mersenne, probably the leading physicists of the time. And if peoples' miseducation in science has simply taught them to be dogmatists, they can hardly complain if those on the outside can only see dogmatism. Whereas the reality is that science is a human activity, not an abstract calculus, and this properly makes its great achievements a subject of pride and awe, not suspicion and scepticism. And it should also make us aware of its desperate fragility, and the hostile cultural forces that it constantly has to overcome.

III

Many writers accept a version of what has become known as the "no miracles" argument for science's claim to depict reality truly. This starts with some uncontested fact about the success of a science, such as its accuracy of prediction, or its technological application. Our lasers and cellphones work, our materials have their calculated strengths, our predictions are borne out to extraordinary numbers of decimal places. What can explain this except that we are getting things right, or very nearly right, or in other words, that we are on the track of the truth? If we were not, it would be an inexplicable coincidence, a miracle, that we are so often so successful.

The argument is compelling, and I completely accept it. But we need to wonder what it is about truth that makes it compelling. Let us take any instance of scientific success. A GPS receiver tells you where you are with astonishing accuracy, based on its distance from four or more satellites orbiting the earth. How does it know those distances? It uses a time differential and the speed of light, and for simplicity let us consider only the speed of light. What then explains the instrument's accuracy? Science says that the speed of light is so many metres per second, and that's the correct, or the true value. It is the truth of the estimate that is vital to the working—if we had got it wrong, and not by much, the instrument would be useless.

Here truth is in the shop window, as it were. But the curious thing is that we can put the identical explanation without mentioning truth at all. Pick up the story right at the end: what explains the instrument's accuracy? Science says that the speed of light is so many metres per second, and that's true. Or, science says that the speed of light is so many metres per second and the speed of light is so many metres per second. The second makes no mention of truth, but it works just as well to explain our success. Indeed it has some title to being science's own explanation of it, and that is the best that there is. Science does not typically mention the concept of truth in describing how GPS devices work.

It is a queer thing about truth that it has this self-effacing quality. And it is not as if we have to choose which of the explanations should be preferred, the one with truth in the shop window or the one without. They come to exactly the same thing. Many philosophers, myself included, think that this implies that the notion has a logical, rather than a metaphysical, function. A large claim such as 'Science gives us the truth' would be a summary way of collecting together a lot of examples such as 'Science says that cholera is due to a bacterium, and it is' and 'science says that the earth circles the sun, and it does'. Since we all assent to many, many, such examples, we can summarize our confidence by assenting to the generalization as well.

If truth retires into the shadows as an interesting topic, so do its detractors. Rorty's campaign, for instance, evaporates because whether there were once dinosaurs is one thing; whether our peers let

us get away with saying it is patently something else. But evidence can occupy some of the vacuum left by any more substantive conception of truth. The problem with flat earthers, creationists, homeopaths and the rest is then not so much that they have a duff conception of truth, as that they have duff attitudes to evidence. The problem with creationists, for example, is that they either know nothing about either stratigraphical or radiometric dating of geological time, or they misunderstand them, or at the worst they have some fanciful notion that uniformities in nature are not the things to rely upon, in which case they might as well believe that they themselves and their sacred books were all created at the same time, say a couple of minutes ago. If we cannot take what is uniformly the case within our experience as our guide for hypotheses about regions of the world beyond it, then reasons dissolve and all bets are off. Reliance on such regularity, as Hume saw, is necessary if we are to move one step beyond the immediately given, and in fact, as Kant added, it is necessary in order to think of ourselves as inhabiting a world at all. It is a necessary presupposition of thought itself. So when the creationist arbitrarily strays from relying on regularities, he has to be betraying the very reasoning that he himself constantly uses.

So once again we return to the problem of induction. The word 'faith' is going to raise its annoying head at this point. Is the human reliance on uniformities just as much a matter of faith as the Creationist's reliance on whatever message tells him that the earth is six thousand years old? A lot of modern writing in the theory of knowledge more or less throws in the towel and supposes that it is. Wittgenstein summed it up in his last book, *On Certainty*, arguing that what we would like are rock-solid foundations for our beliefs, but what we find are things that simply "stand fast" for us—and that, of course, raises the disturbing possibility of others for whom different and in our eyes deplorable things equally stand fast. This is really only a rediscovery of Hume's own results. But faith is the wrong word, if it implies cousinship with arbitrary stabs of confidence in things for which there is no evidence. Those can be avoided, and should be. Whereas, as I argued above, Hume and Kant show that a modest confidence in the wonderful stabilities of the world goes with our capacity to think at all.

The history and philosophy of science provide the most important defences our culture has. They can only be defended if we have a proper understanding of their epistemology. Although Popper ranks high as someone who has worked in just that field, I fear we have to stand on his shoulders, and do better if we can.

Justification in Historical Natural Science

Carol E. Cleland[♣♦]

Introduction

Historical research is common in natural science, occurring in fields as diverse as paleontology, geology, biology, planetary science, astronomy, and astrophysics. Much of this work involves explaining puzzling contemporary phenomena (traces) discovered through fieldwork in terms of conjectured, long-past causes. In recent years the historical natural sciences have enjoyed an increasing number of high profile successes. Some celebrated examples are: the hypothesis that the continents were once joined together into a super continent (Pangaea), which explains surprising patterns of frozen magnetism found in certain ancient igneous rocks; the Alvarez meteorite-impact hypothesis, which explains the startlingly high concentrations of iridium and shocked quartz found in the mysterious K-T (Cretaceous-Tertiary) boundary marking the end of the fossil record of the dinosaurs; and the big-bang theory of the origin of the universe, which explains the mysterious isotropic, 3° Kelvin, background radiation first detected by satellites in the mid-1960s. Yet with the exception of evolutionary biology, philosophers of science have displayed little interest in the historical natural sciences. This is particularly puzzling when one considers that the methodology of historical natural science does not seem to closely resemble that of stereotypical experimental science, the latter of which is commonly held up as the paradigm of "good" science.

I addressed this issue in earlier work (Cleland [2000], [2001]), identifying fundamental differences in the practices of "prototypical historical natural science" and "classical [stereotypical] experimental science." Unlike the hypotheses of classical experimental science, which postulate regularities among types of events, the hypotheses of prototypical historical science are concerned with long-past, particular events, e.g., a specific meteorite impact as opposed to meteorite impacts in general. The acceptance and rejection of hypotheses in classical experimental science depends upon the success or failure of predictions tested in controlled laboratory settings. In contrast, the acceptance and rejection of hypotheses in prototypical historical natural science depends upon their capacities to *explain* puzzling collections of traces discovered through fieldwork. I argued that these differences in practice could be explained in terms of a physically pervasive time asymmetry of causation, dubbed the "asymmetry of overdetermination" by philosopher David Lewis ([1979]). The asymmetry of overdetermination underpins the objectivity and rationality of the practices of prototypical historical natural science, explaining why it is not, as sometimes supposed, inferior to experimental science.

This essay explores the structure of historical explanation and its epistemic role in confirming and disconfirming conjectures about long-past, particular events; for more detail, see Cleland ([2009], [forthcoming]). I begin, in section 2, by reviewing my earlier analysis of the methodology of historical natural science. In section 3, I argue that the acceptance and rejection of historical hypotheses in the natural sciences cannot be understood in terms of their predictive capacities. Explanation in historical natural science is grounded in causal considerations. The most fundamental mode of causal explanation is common cause explanation; common cause explanations supply the evidential warrant for all conjectures in the natural sciences concerning particular (vs. generic) long-past events. As discussed in Section 4, common cause explanation is traditionally justified by appealing to the principle of the common cause. But what justifies the principle of the common cause? Some philosophers have argued that it is highly problematic because it is either purely methodological or strictly metaphysical. I argue that the principle of the common cause is empirically well grounded in the asymmetry of overdetermination, a physically pervasive time asymmetry of causation, as opposed to logic or metaphysics. Conceptualizing the principle of the common cause in terms of the asymmetry of overdetermination helps to explain some otherwise puzzling characteristics of the methodology of historical natural science.

♣ Philosophy Department, Center for Astrobiology, University of Colorado, Boulder, CO

♦ This work was supported in part by a NASA astrobiology grant to the University of Colorado's Astrobiology Center.

The Methodology of Historical Natural Science

In earlier work (Cleland [2001], [2002]), I argued that most (prototypical) historical research in natural science exhibits a distinctive pattern of evidential reasoning characterized by two interrelated stages (1) the proliferation of multiple, competing, alternative hypotheses to explain a puzzling body of traces encountered in fieldwork, and (2) a search for a "smoking gun" to discriminate among them. A smoking gun discriminates among rival historical hypotheses by showing that one or more provides a better explanation for the total body of evidence available than the others. As I emphasized ([2002]), this pattern of evidential reasoning is not always found in the historical natural sciences, and it is sometimes found in (non-classical) experimental research. Which pattern of evidential reasoning is exhibited depends upon a scientist's epistemic situation.

The stages that I identified in prototypical historical natural science are not, as Kleinhans et al. ([2005]) assert, in conflict. The body of evidence on the basis of which a collection of rival hypotheses is formulated does not include the smoking gun that subsequently discriminates among them. A smoking gun represents a piece of additional evidence that wasn't available at the time the hypotheses concerned were formulated; undiscovered traces do not constitute actual evidence. The discovery of a smoking gun changes the evidential situation, revealing that one or more of the hypotheses under consideration provide a better explanation for the total body of evidence *now* available than the others. Furthermore, an investigation may be quite dynamic. The original collection of competing hypotheses may be culled and augmented repeatedly in light of new evidence and/or advances in theoretical understanding. Ideally this process converges upon a single hypothesis. But there are no guarantees. And even supposing that a scientific consensus is reached on a single hypotheses, there are no guarantees that future empirical or theoretical work won't bring to light scientifically viable, new possibilities. If this happens, the previously well-accepted hypothesis will acquire a rival, and the process of searching for a smoking gun begins anew.

In this context, it is important to keep in mind that there isn't a guarantee that the correct hypothesis is among those being entertained, or for that matter, that it will ever be entertained by humans; historical scientists are just as limited by their imaginations as experimentalists. Besides, even supposing that the correct explanation is among those under consideration, there are no guarantees that a smoking gun for it will be found even supposing that one exists. Breakthroughs in historical science frequently wait upon the development of sophisticated technologies for detecting and analyzing miniscule or highly degraded traces. In the absence of the requisite technology, historical scientists have little choice but to resign themselves to a collection of equally viable, rival hypotheses.

A Case Study: The Alvarez (Meteorite-Impact) Hypothesis

The scientific debate over the end-Cretaceous mass extinction, which famously killed off the dinosaurs, along with what is now estimated to be 75% to 85% of all species then on Earth, provides a particularly good illustration of the dynamic interrelation between proliferating alternative hypotheses and searching for a smoking gun to discriminate among them. Prior to 1980, many different explanations were taken seriously by paleontologists, including pandemic, evolutionary senescence, climate change, nearby supernova, volcanism, and meteorite impact (Powell [1998], p. 165). Most of these hypotheses explained the fossil record of the dinosaurs by postulating mutually incompatible common causes. None of the evidence available at the time, however, provided strong support for any one of these hypotheses over the others, and most paleontologists suspected that we would never know which is correct. It thus came as a surprise when the father and son team of Luis and Walter Alvarez ([1980]) discovered something momentous in the K-T boundary.

Found all over the world, the K-T boundary marks the end of the Cretaceous and the beginning of the Tertiary (the "age of mammals"). It consists of a very distinctive, thin layer of clay sandwiched between two layers of limestone, suggesting a sudden collapse of biological activity. Geologists long suspected that it held the secret to the end-Cretaceous mass extinction, but no one knew how to unlock it. Walter Alvarez, a geologist, was interested in how long it took for the K-T boundary sediments to be deposited; was the extinction event rapid or slow? His father Luis, a

physicist, suggested using the element iridium as a clock since it is supplied at a known constant rate by meteoritic dust. Detecting the expected low levels of iridium required a nuclear reactor (particle accelerator), which Luis had access to at Berkeley; when bombarded with energetic neutrons, an isotope of iridium emits distinctive gamma rays, allowing the amount of iridium in the sample to be determined by counting the number of rays. The results were staggering. Clays from the K-T boundary contained iridium levels 30 times higher than the limestones on either side. Luis's calculations showed that the amount of iridium was too great to be explained in terms of known geological processes. Subsequent tests confirmed the presence of an iridium anomaly in K-T boundary clays from around the world.

Luis and Walter knew that they were in possession of a smoking gun for the mysterious end-Cretaceous mass extinction. Earth's crust is depleted in iridium because iridium (like iron) is a heavy element and most of it sank into the mantle and core during planet formation. Although not all meteorites are rich in iridium, asteroids and comets left over from the formation of the solar system typically have higher concentrations. So meteorite-impact was a very promising candidate for explaining the anomalous levels of iridium. On the other hand, as volcanologists (e.g., Officer and Drake [1985]) pointed out, volcanism brings mantle material to the surface. Moreover, there is evidence of extensive flood volcanism (spread over an area of at least 1 million km^2) in the Deccan traps region of India approximately 65 mya (million years ago). Accordingly, volcanism provides an alternative possibility for explaining the iridium anomaly in K-T boundary sediments. None of the other competing hypotheses for the end-Cretaceous mass extinction could explain the excess iridium. The Alvarezes' discovery of anomalous levels of iridium in the K-T boundary thus functioned as a smoking gun for discriminating meteorite impact and volcanism from their pre-1980 rivals.

Further research supported meteorite impact over volcanism. Fieldwork undercut the claim that volcanism could produce a global iridium anomaly (e.g., Schmitz and Asaro [1996]). More importantly, however, further analysis of K-T boundary sediments produced a smoking gun for meteorite impact over volcanism. Large quantities of mineral grain, predominately quartz, exhibiting a highly unusual pattern (crosshatched, parallel sets) of fractures was found in K-T boundary sediments from around the world (Bohor et al. [1984]). It takes enormous pressures to fracture minerals in this way. There were only two places on Earth where they were known to occur, the sites of nuclear explosions and meteor craters. Subsequent fieldwork failed to substantiate the claim that extremely violent volcanic eruptions produce minerals of this sort (Kerr [1987]; Alexopoulos et al. [1988]). The combination of excess iridium and shocked quartz in the K-T boundary was thus enough to convince most members of the scientific community that a large (10-15 km wide) meteorite hit Earth 65 million years ago. Since this time more evidence of meteorite impact (microspherules, fullerenes containing extraterrestrial noble gases, and extensive deposits of soot and ash) has been discovered in the K-T boundary. But it is generally agreed by planetary and earth scientists that the combination of an iridium anomaly and shocked minerals cinched the case early on. [8]

---

[8][8] In this context it is worth noting that although the discovery of the Chicxulub crater, which is roughly 200 km across and straddles the northern coast of Mexico's Yucatan Peninsula, is sometimes cited as pivotal, it was not. It is difficult to connect even a gigantic, local impact crater with a global extinction. In contrast, the global distribution of iridium and shocked quartz in K-T boundary sediments from around the world points to a meteorite impact with global, and hence potentially catastrophic, effects. Had the Chicxulub crater been discovered in the absence of the iridium and shocked quartz, it is unlikely that it would have been construed as compelling evidence for a meteorite-impact explanation for the end-Cretaceous extinctions. On the other hand, once the iridium and shocked quartz were discovered, it wouldn't have surprised scientists if no one had been able to locate a crater of the right size and age since seventy percent of Earth's surface is covered by ocean, making an ocean impact more probable than a land impact, and an ocean impact crater would almost certainly have been obliterated by now the active geology of the seafloor, which moves in conveyor like fashion away from mid-ocean ridges, where it forms, to the margins of continents, where it sinks back into the mantle at subduction zones. Indeed, many geologists, who were convinced by the iridium and shocked quartz that a devastating meteorite impact occurred around 65 million years ago, were pleasantly surprised when a crater of the right size and age was identified straddling a landmass; they didn't view the case for a catastrophic meteorite impact as resting upon the discovery of an appropriate crater. This is not to deny that many scientists view the discovery of the crater as augmenting an initially strong case. Significantly, however, a few geologists, who are also convinced by the iridium and shocked quartz that a catastrophic impact occurred 65 million years ago, insist that the Chicxulub crater isn't the right one.

The iridium and shocked minerals weren't enough, however, to convince most paleontologists that the second prong of the Alvarez hypothesis is true—that the impact caused the mass extinction. The extinctions had to be worldwide and geologically instantaneous. The available fossil evidence was very imprecise, unable to distinguish extinction events occurring within a period of a few years from those occurring at different times throughout intervals of 10,000 to perhaps 500,000 years. Moreover, some of the fossil evidence seemed to suggest that the extinctions were well underway by the time the impact occurred (Clemens et al. [1981]), leading some paleontologists to infer that something else (climate change, evolutionary senescence, or extensive volcanism were some popular conjectures) was at fault, and the impact, at best, delivered the *coup de grace*. Additional fieldwork was needed to establish a more convincing causal link between the impact event and the extinction event.

Paleontologists went to work, closely studying the fossil records of different kinds of organisms on either side of the K-T boundary. Peter Ward ([1990]) established that the fossil record of the ammonites goes right up to the K-T boundary and then suddenly disappears. Studies also documented substantial changes in the morphology of the calcareous shells of tiny planktonic foraminifera on either side of the K-T boundary. Paleobotanists made some of the most significant fossil discoveries. Using high-resolution techniques, they discovered abundant fossilized angiosperm (flowering plant) pollen right up to the lower level of the boundary, at which point it disappears and is replaced, on the other side of the boundary, with abundant fossilized fern spores (Johnson and Hickey [1990]). As botanists know from experience with modern catastrophes (e.g., the explosion of Mount St. Helens) ferns are opportunistic plants that quickly colonize devastated areas. These detailed fossil studies from around the world indicated that the extinction was massive (involving many different kinds of organisms), rapid, and catastrophic. Most paleontologists were won over to the second prong of the Alvarez hypothesis, illustrating that a smoking gun may consist of a large and diverse body of new evidence.

The remarkable cross-disciplinary, scientific consensus that was finally achieved on the Alvarez hypothesis stands as one of the crowning achievements of historical natural science. As a consequence it provides a particularly compelling case study for evaluating philosophical theories of historical natural science. For this reason I appeal to it extensively in subsequent discussions.

Justification in Historical Natural Science

Unlike confirmation in classical experimental science, which is grounded in prediction, confirmation in prototypical historical science depends upon explanatory power. The iridium anomaly, which played such a pivotal role in the acceptance of the first prong of the Alvarez hypothesis, provides a salient example. The Alvarezes didn't predict excess iridium in the K-T boundary, and then set out to find it. They literally stumbled upon it while exploring a different question: How long did it take for the boundary layer to be deposited? The significance of the iridium anomaly for the Alvarez hypothesis lies in the fact that in the context of the scientific knowledge available at the time the latter (with the possible exception of the volcanism hypothesis) provides a better explanation for the former than any of the competing hypotheses.

It is important to appreciate that no one could have predicted (in the sense of logically inferred) an iridium anomaly from the conjecture that a gigantic meteorite struck Earth 65 mya on the basis of the scientific knowledge available at the time of the Alvarezes' discovery. Furthermore, even today there aren't any widely accepted, background auxiliary hypotheses that could warrant such an inference. Our current understanding of earth and planetary science informs us that there are many highly plausible, extenuating circumstances capable of defeating an inference to an iridium anomaly from a gigantic meteorite impact, e.g., an iridium-poor meteorite, dispersal of an initial iridium anomaly by geological processes, and unrepresentative samples of the K-T boundary. Peter Ward's pivotal studies of Cretaceous ammonites provide a good illustration of the threat posed by unrepresentative samples. Exposed outcrops of the K-T boundary are very rare, many are still buried, and of those that have been exposed, the majority has long since been removed by erosion. Ward was working on the Spanish side of the Bay of Biscay, whose sea cliffs contain abundant ammonites and some of the best exposed, well-preserved outcrops of the geological section containing the K-T

boundary in the world. The closest ammonite that he could find to the lower level of the boundary was 10 meters beneath it, leading him to suspect that they had become extinct tens of thousands of years earlier (Ward [1983]). Serendipitous (as it turned out!) encounters with armed Spanish soldiers and disgruntled Basques eventually motivated him to change location, and he moved a short distance up the coast to France, where, to his surprise, he found abundant ammonites extending right up to the boundary. Apparently the ammonites in what is now northern Spain suffered an ecological crisis during the late Cretaceous but continued to thrive just a few miles up the coast, in what is now southern France. Ward ([1990]) concluded that the fossil record of the ammonites supported (the second prong of) the Alvarez hypothesis after all!

Ward nonetheless characterized his fieldwork as testing a "prediction" of the Alvarez hypothesis. The question is what sort of a prediction could it be? As the discussion above makes clear, it cannot be a *precise* prediction to the effect that ammonite fossils will be found extending right up to the K-T boundary on the Spanish side of the Bay of Biscay. At best, it may be interpreted as a vague prediction to the effect that it is *likely* that there are rock sequences *somewhere* on Earth with ammonite fossils immediately below the lower edge of the boundary. Viewed from this perspective, Turner ([2007], Ch. 5) is correct when he says that historical scientists sometimes infer novel predictions from hypotheses about the past. But the fact that they do so does not, as he suggests, show that prediction plays the same role in prototypical historical science as it does in classical experimental science. The problem with vague prognostications like Ward's is that they are virtually immune to failure. Failure to find ammonite fossils in any particular location or in a number of particular locations doesn't bear on the possibility of finding them elsewhere, in some as yet unexplored rock record of the K-T boundary. This would be true even if he had failed to find them on the French side of the Bay of Biscay. This is in contrast to classical experimental science where failed predictions have as much if not more weight than successful predictions.

Hypotheses in classical experimental science are concerned with regularities, as opposed to singular events. Moreover the spatio-temporal gap between cause and effect is small. As a consequence, they can be repeatedly "tested" in localized laboratory settings by varying certain conditions while holding others constant. This makes it easier to identify or eliminate potential interfering conditions, and thus more difficult (but *a la* Duhem not impossible) to explain away failed predictions. The situation in prototypical historical natural science is quite different. Historical hypotheses are concerned with singular occurrences and the causal chain extending from cause to present-day effects is typically extremely long, complex, and convoluted. This makes prediction a poor tool for guiding a search for telling empirical evidence. Historical scientists know that there are many poorly understood (including unknown) background conditions and potentially interfering factors that could defeat a prediction independently of the truth of the hypothesis, and they don't have a good way of screening for which ones might have been operative in a given case. The upshot is that novel predictions in historical science are typically much less risky (in Popper's sense) than those in classical experimental science.

Somewhat ironically, this point is underscored by the examples of failed novel predictions cited by Turner ([2007], Ch. 5). The snowball Earth hypothesis provides a salient illustration. According to Turner, the snowball Earth hypothesis holds that the entire planet was completely covered in ice for several million years on several different occasions during the neoproterozoic (ca. 850-555 mya). Some physical geologists suspect that an event this extreme would produce a planet-wide "hydrological shutdown," which provides the basis for a novel prediction: geological sections of the pertinent age should reveal periods during which no sediments were formed (because no weathering occurred). Leather and colleagues (2002) set out to test this "prediction" in northern Oman, which is one of the few places on Earth where one can find neoproterozoic deposits of the right age. But they didn't find evidence of a hydrological shutdown. They discovered bands of glacial debris interspersed with and broken up by layers of sediment deposited over a fairly short time period.

The paleogeological community did not, however, respond to Leather and colleagues' discovery by rejecting the snowball Earth hypothesis, and for very good reasons. First, the snowball Earth hypothesis was never as specific as Turner suggests. From the beginning, there was disagreement about whether the planet was almost or completely covered in ice (were there any areas

of open ocean?), how hard the freeze was (slushy or frozen solid at the equator?), how long individual episodes lasted, etc. Second, the claim that a snowball Earth would produce a planet-wide hydrological shutdown, in which no sedimentary (including glacial) deposits are formed for a long period of time, was based upon climate models incorporating a large number of somewhat speculative background assumptions about atmospheric, oceanic, and continental conditions and processes, e.g., atmospheric CO_2 levels, extent of sublimation processes over sea-ice, thickness of sea-ice cover, thickness of continental ice sheets, varieties of non-hydrological processes of chemical weathering, length of total freezes, paleoaltitude and tectonic evolution of the continents, frequency of interglacial/nonglacial periods, etc. As a consequence, the claim that no sediments would be deposited during a snowball Earth episode was open to question. Third, there was the problem of interpreting what is found at a unique geological site; subsequent geological processes may intermingle material deposited at different times, producing misleading rock records and radiometric ages. Given these background considerations, it should come as no surprise that the debate over the snowball Earth hypothesis continues to this day, with some researchers (see Fielding et al. [2006]) contending that although the glaciations were nearly planet-wide, they were of short duration, alternating with longer periods of warmer, interglacial conditions, and that sublimation of sea-ice drove a significantly diminished (but not fully shut down) water cycle.

Predictions that succeed, in contrast, sometimes carry great weight in prototypical historical natural science. But it is not in virtue of representing a successful novel prediction that they do so. Regardless of the circumstances in which it is acquired, evidence functions as a smoking gun if it shows that one hypothesis provides a better explanation than its rivals. If Ward had accidentally stumbled upon ammonite fossils just below the K-T boundary in France, as opposed to having gone looking for them there, his finding wouldn't have been any less significant. This explains why so many of the high profile achievements of historical science have the character of serendipitous discoveries even when they can be interpreted as involving novel predictive successes. In this context it is important to keep in mind that the evidence that makes a vague prediction successful may itself be quite precise. Ward's discovery in France was not vague: He found abundant ammonites within a meter of the lower edge of the K-T boundary in a well-preserved outcrop of the pertinent geological section. This discovery provides much better evidence for the conjecture that the ammonites did not go extinct before the impact than his failure to find ammonites in an analogous rock record in Spain provides evidence that they went extinct.

As Turner admits, cases in which historical hypotheses are rejected on the basis of failed predictions are the exception rather than the rule. He pins the problem on the difficulty of "testing" novel predictions in the historical sciences. In so doing, he implicitly endorses the widely accepted assumption that the practices of stereotypical experimental science provide the prototype for all of science. It is thus hardly surprising that he concludes that historical science is epistemically disadvantaged vis-à-vis experimental science (Turner [2004], [2007]). But as I have argued ([2001], [2002]), the actual practices of historical natural scientists provide little support for this assumption. Most historical hypotheses are not rejected on the basis of failed predictions but rather because another hypothesis does a much better job of explaining the total body of evidence available in the context of our scientific background knowledge. As an example, the contagion hypothesis for the extinction of the dinosaurs cannot be viewed as refuted by the discovery of the iridium anomaly because, as the scientists involved would readily admit, the presence of iridium in the context of their background understanding of Earth history does not provide evidence that the dinosaurs did not go extinct as a result of an epidemic shortly before or after the impact. What the discovery of iridium did was to provide positive support in the form of independent evidence for either a gigantic meteorite impact or massive volcanism, either of which has the capacity (under the right circumstances) to produce a mass extinction. It is thus not an accident that scientists did not speak of the contagion hypothesis as being "refuted" by the discovery of iridium in the K-T boundary. Instead they simply stopped talking about the contagion hypothesis and moved on to the question of whether extensive flood volcanism or a gigantic meteorite impact provides the best explanation for the iridium anomaly. The point is in historical science a hypothesis may be rejected on the basis of evidence that does not refute (or falsify) it.

In sum, unlike classical experimental science, prototypical historical natural science is not a prediction-centered enterprise. Hypotheses about long-past, particular events are "confirmed" by evidence in virtue of their power to explain, as opposed to successfully predict, the evidence. This is not just a matter of *accommodation*. The evidence (smoking gun) that cinches the case for an historical hypothesis over its rivals is typically discovered after the hypothesis was formulated. The Alvarez hypothesis for the end-Cretaceous extinctions provides a good example. The hypothesis did not originate with the Alvarezes, despite the fact that it now bears their name. It was propelled from the backburner to the frontburner of geological science with the discovery of positive evidence that such an event actually happened. Nor can the acceptance of the Alvarez hypotheses be construed as a matter of *retrodiction*: contemporary scientists do not have the requisite background knowledge to *logically* infer a meteorite impact from the discovery of an iridium anomaly any more than they do to logically infer an iridium anomaly from a conjectured meteorite impact. Geological processes are just as capable of concentrating material that was originally dispersed as they are of dispersing material that was once concentrated. A good example is placer deposits, which are formed when flowing water picks up weathered minerals from different regions and eventually deposits them together (sorted according to weight) in an area of less rapid flow. Half of all gold ever discovered, for instance, comes from placer deposits. Placer deposits are just one of several geological processes that can concentrate dispersed minerals.

A scientific consensus on the meteorite-impact hypothesis for the K-T extinctions was achieved because it *explains* an otherwise puzzling body of traces, many of which (e.g., iridium, shocked quartz, glassy spherules, etc., and fossil records of ammonites, foraminifera, plant pollen, fern spores, etc.) were discovered after the hypothesis had been formulated, better than any of its competitors. The appearance of these disparate traces in geological strata of the same age is deeply mysterious; they are individually unexpected and their joint occurrence is even more enigmatic. The Alvarez hypothesis explains this double mystery better than any of its currently available, scientifically plausible competitors. It is for this reason that it is currently widely accepted by the scientific community.

The Structure of Historical Explanation

At one time the emphasis on explanation over prediction in the confirmation of historical hypotheses wouldn't have been viewed as significant. For on the traditional covering law model of scientific explanation, prediction and explanation have the same logical structure (Hempel and Oppenheim [1948], Hempel [1965]). The prototype for the covering law model, the D-N (deductive-nomological) model, analyzes explanations as deductively valid arguments whose premises are statements of general law and (sometimes but not always) initial conditions, and whose conclusions are statements of the phenomenon (event, fact, or regularity) to be explained. Every adequate explanation is thus a potential prediction (Hempel [1965], p. 367). In order to accommodate statistical or probabilistic laws, Hempel augmented the covering law model with the D-S (deductive statistical) and I-S (inductive statistical) models of explanation; Hempel assumed that there are logical principles of inductive inference analogous to those of deductive inference. All three models analyze explanations as arguments in which the explanatory burden rests upon laws of nature.

Historical explanation was a problem for the covering law model from its inception. Laws (whether deterministic or statistical/probabilistic) that are strong enough to license logically 'valid' deductive or inductive inferences must be universal (within the pertinent domain of discourse) and exceptionless. Explanations in historical science rarely invoke even rough generalizations of this sort. The long causal chain stretching between a prehistoric event and its contemporary traces is just too complex, involving the intersection of many independent causal chains, to be captured in a plausible generalization of the kind required by the covering law model; compelling statistical or probabilistic laws require reliable information about frequencies, which is rarely available, particularly in cases involving uncommon events such as mass extinctions. Hempel was fully aware of this difficulty. His solution was to demote historical explanations to mere "explanatory sketches", thus reinforcing the widely accepted view that the historical natural sciences are inferior to the experimental sciences. Hempel attributed the undeniably compelling nature of some historical explanations to the tacit assumption of unspecified natural laws, a view which still attracts adherents, e.g., Ereshefsky [1992]).

But this represents little more than an *ad hoc* attempt to force historical explanations to conform to a favoured but inadequate model of scientific explanation.

Common strategies for dealing with the restricted, exception-ridden generalizations of the "special sciences" are to tack on *ceterus paribus* clauses. As Sandra Mitchell ([2000], [2002]) argues, however, this strategy is not very satisfactory. To be compelling *ceterus paribus* laws require approximate generalizations coupled with knowledge of some contingencies (interfering factors); the *ceterus paribus* clause magically absorbs any additional, unknown dependencies. But even approximate generalizations are strikingly absent from most explanations in historical science. Scientists just don't know enough about all the things that might happen in the spatio-temporally extended causal chain linking a postulated long-past cause to its present day traces to determine what should be included in an approximate generalization and what should be consigned to a *ceterus paribus* clause.

Kleinhans and colleagues ([2005]) embrace a reductionist solution to this difficulty. On their view, the extremely rough generalizations of contemporary geology (historical and nonhistorical) are "reducible" to the stricter generalizations of chemistry and physics. In their words, "earth science generalizations, such as the cited example regarding earthquakes, describe contingent distributions and processes which can be reduced 'locally' because they can be *exhaustively* [italics are mine] translated in physical and/or chemical terms" (p. 295). But what evidence (other than blind faith) is there for this? Geologists are notoriously bad at predicting earthquakes even for extensively studied, local regions of well-mapped fault systems such as the San Andreas Fault. Moreover, even supposing that it is *in principle* possible to reduce generalizations distinctive of earth science to laws of physics and chemistry, no human being knows how to do the reductions. This means that the conjectured reducing laws do not play an actual role in explanations given by contemporary earth scientists. Besides, as Nancy Cartwright (e.g., [1983]) has argued, it isn't even clear that the laws of fundamental physics are universal and exceptionless. Kleinhan's and colleagues' proposal amounts to little more than a return to Hempel's faith-based explanatory sketches.

In light of these and other considerations, Mitchell proposes modifying the traditional concept of law of nature to include degrees of contingency or, in her words, "stability over changes in context" (Mitchell [2002], p. 334). The laws of fundamental physics exhibit the greatest (but not perfect) stability and the laws of the special sciences the least. In this way she hopes to preserve the central role of laws of nature in scientific reasoning. Ben Jeffares (2008) embraces Mitchell's concept of law of nature and argues that the investigation of rough generalizations is just as central to prototypical historical natural science as is the search for a smoking gun. Like Hempel and fellow travelers, Jeffares is convinced that the ultimate source of evidential warrant for scientific hypotheses lies in the success or failure of predictions. In his words, "the historical sciences also seek regularities in the world and *have to* [italics are mine] in order to secure their claims about the past" (p. 470).

There is little doubt that historical scientists deploy generalizations from the experimental sciences in analyzing and interpreting traces discovered in the field. A salient example is the use of radiometric dating methods, which are grounded in the highly stable, statistical laws of quantum theory. It is clear, however, that generalizations of this sort play a secondary role in historical research. They are not the targets of historical research but rather useful tools borrowed from other disciplines for special purposes. It is also true that historical scientists sometimes investigate much less stable, special-purpose regularities in laboratory settings. Jeffares cites archaeologists "experimenting" with differences in marks produced by dogs gnawing bones and humans using primitive tools to butcher animals as an example. It is clear, however, that this regularity is being pursued as a means to an end, as opposed to "an end in itself" (p. 470). As Jeffares concedes, archaeologists are interested in discriminating marks on bones left by human tools from those left by canine teeth for the purpose of interpreting marks found on ancient bones; the purpose of the experimental work is to procure a tool (analogous to radiometric dating methods) for analyzing evidential traces discovered in the field.

The question is whether Jeffares is correct in claiming that historical scientists investigate special primary generalizations—generalizations holding "directly" between long-past causes and their contemporary traces—and utilize them for purposes of prediction. As the discussion in the preceding section underscores, this is a very problematic claim. Regularities holding between causes

and effects separated by protracted intervals of time are exceedingly fragile; each link in the causal chain represents a causal liability (an opportunity for interference), and the longer the time span, the greater the number of contingencies that the generalization must accommodate. The upshot is that it is not only difficult to identify generalizations holding directly between long-past causes to their present day traces, any generalizations that are identified are extremely unstable (in Mitchell's sense). One cannot infer predictions capable of playing pivotal roles in the evaluation of hypotheses from generalizations with this degree of contingency. Jeffares's mistake is in thinking that he can retain the explanatory power of prediction (*à la* the covering law model) with a much weaker notion of natural law.

The purpose of this discussion has been to establish that it is not in virtue of functioning as (successful or failed) potential predictions that explanations "confirm" and "disconfirm" hypotheses about long-past particular events. At one time this would have been considered grounds for thinking that historical science is inferior to experimental science. But the covering law model has fallen on hard times in recent years even for explanations in physics, and attempts to fix it up have not been very successful. The dominant contemporary philosophical theories of scientific explanation place the explanatory burden on causal features of the world. In keeping with some contemporary metaphysical accounts of causation, many causal theories of explanation are open to the possibility that some potentially explanatory causal relations do not come under natural laws of any sort. In short, causation is thought to be more essential to causation than lawfulness.

The Centrality of Common Cause Explanation

The two main modes of causal explanation in the historical natural sciences are *narrative explanation* and *common cause explanation*. Narrative explanation dominates thought about explanation in human history, where intangible human desires and purposes play key explanatory roles. It is also common in evolutionary biology and historical geology. The basic idea behind narrative explanation is to construct a "story"—a coherent, intuitively continuous, causal sequence of events centered on the event to be explained. Because much is unknown about the events in the sequence, narrative explanations have a significant fictional component, involving both omissions and additions. This poses a potential problem insofar as it conflicts with the traditional emphasis in natural science on evidential warrant. The problem is exacerbated by the central role of explanation in the confirmation of historical hypotheses. If the primary reason for accepting a historical hypothesis is its explanatory power and it draws its explanatory power primarily from the coherence and continuity of a quasi-fictional story, then historical natural science really does seem inferior to experimental science.

Common cause explanation promises a solution to the problem of evidential warrant posed by narrative explanation. The basic idea behind common cause explanation is to formulate reliable inferential methods for identifying when a diversity of contemporary traces comprises the effects of a long-past, common cause token. It is thus not surprising that narrative explanations and common cause explanations frequently go hand-in-hand in historical natural science (e.g., Ruse ([1971]) and Hull ([1992])), with common cause explanations supplying the needed epistemic support for key events in the narrative sequence. The increasingly detailed narrative for the end-Cretaceous mass extinction provides a salient illustration. The discovery of large quantities of rain-drop shaped, glassy spherules and extensive deposits of soot and ash in K-T boundary sediments from around the world, for instance, supports the claim that enormous quantities of rock were liquefied and vaporized during the impact (including the entire meteorite) and injected into the upper atmosphere only to fall back, after enveloping the planet, in a global "rain" of fire, igniting everything that could burn on the planet's surface. Another good illustration is provided by the phylogenies (evolutionary histories) constructed by biologists for organisms and groups of organisms. The discovery of "molecular fossils" in living organisms (genomic sequences that have changed little over the eons) has given a tremendous boost to some phylogenies while discrediting others because they provide new evidence (in addition to that of morphology and the fossil record) of common ancestry.

While narrative explanations derive their empirical support from common cause explanations, not all common cause explanations are associated with narrative explanations. A good illustration is paleontologist Mary Schweitzer and colleagues' ([2005]) explanation for what appears to be medullary

bone inside the fossilized leg bone of a Tyrannosaurus rex. Medullary bone comprises a distinctive calcium rich layer that develops in the long bones of contemporary female birds during the egg laying process, providing a readily accessible supply of calcium for building eggshells. Schweitzer and her graduate student were stunned when they discovered an analogous layer in the fossilized leg bone of a T. rex. They concluded that the bone was from a female T. rex. Significantly, they did not concern themselves with the circumstances of the death of this unfortunate T. rex, nor did they attempt to reconstruct any of the events in the long causal chain stretching between its death and the preservation of its bone for millions of years in the Montana desert. Indeed, detailed stories of either sort are irrelevant to their purpose, which was simply to evaluate whether the fossilized bone under investigation came from a female T. rex. To this end, they studied the detailed physical structure and chemical composition of the T. rex bone, comparing it to the leg bones of modern female birds and appealing to well-accepted background beliefs about the close phylogenetic relationship between modern birds and dinosaurs. The point is the explanation they gave for the medullary-like bone did not constitute a narrative and was not used to buttress an event in a narrative sequence.

The evidential warrant for hypotheses about long-past, particular events thus ultimately rests upon common cause explanation. Common cause explanation has long been justified in terms of the principle of the common cause, which is traditionally associated with the writings of Hans Reichenbach ([1956]). The version that I defend in this paper is weaker, however, than Reichenbach's. It asserts that seemingly improbable coincidences (correlations or similarities among events or states) are best (vs. must be) explained by reference to a shared common cause.

The principle of the common cause represents an epistemological conjecture about the conditions under which a certain pattern of causation may be non-deductively inferred. According to the principle of the common cause, genuine coincidences are rare. Most coincidences are produced by common causes. Underlying the principle of the common cause is an ostensibly metaphysical claim about the temporal structure of causal relations among events in our universe: most events have multiple effects—form causal forks that open from past to future. In the next section I argue that this presupposition is not merely metaphysical. It is empirically well grounded in physical theory.

The principle of the common cause provides a potentially powerful tool for understanding the close relationship between explanation and confirmation in the reasoning of natural scientists engaged in historical research. Attributing puzzling similarities and correlations among phenomena to a common cause has great explanatory power, for it makes their joint occurrence credible. Attributing their concurrency to chance, on the other hand, explains nothing; we are left with an intractable mystery. The iridium and shocked quartz in the K-T boundary provide a salient example. Given our current understanding of geology, the only event that renders their global concurrence in a structurally distinctive, thin layer of sediment found all over the world explicable is a massive meteorite impact. As a consequence, the case for a meteorite impact is currently overwhelming. Similarly, the best explanation for the truly astonishing structural and chemical similarities between the fossilized leg bone of Schweitzer's T. rex and the long bones of modern female birds is that the former was female. In other words, the more improbable an association among a collection of traces *seems* the more psychologically convincing the claim that it is genuinely improbable, and hence the more compelling (assuming the truth of the principle of the common cause) the claim that it is the product of a common cause.[9] This helps to explain why historical natural scientists have a tendency to focus their investigations on what seems to them (in light of their background beliefs) to be the most puzzling correlations or similarities among contemporary phenomena.

In the following section, I argue that the asymmetry of overdetermination, a physically pervasive time asymmetry of causation, provides the needed objective grounding for the principle of the common cause. It vindicates appeals by historical natural scientists to explanatory successes and failures in their decisions to accept and reject hypotheses about long-past, particular events.

---

[9] Associations that *seem* improbable may not *in fact* be improbable. Whether an association seems improbable depends to a great extent upon the current state of our scientific knowledge. As I will discuss later the discovery of a common cause renders a seemingly improbable event probable; the improbability grows out of our ignorance of the unity of their cause.

The Objectivity and Rationality of Common Cause Explanation

In earlier work (Cleland [2001], [2002]), I argued that the distinctive methodology of prototypical historical natural sciences (proliferating alternative hypotheses and searching for a smoking gun) is best understood in terms of *the asymmetry of overdetermination*. The asymmetry of overdetermination consists in the fact that most local events epistemically *over*determine their past causes (because the latter typically leave extensive and diverse effects) and *under*determine their future effects (because they rarely constitute the total cause of an effect). As an example of the epistemic overdetermination of past causes by their future effects consider an explosive volcanic eruption. Its effects include extensive deposits of ash, pyroclastic debris, masses of andesite or rhyolitic magma, and a large crater. Only a small fraction of this material is required to infer the occurrence of the eruption. Indeed, any one of an enormous number of remarkably small subcollections of effects will do. This helps to explain why geologists can confidently infer the occurrence of long-past events such as the massive, caldera forming, eruption that occurred 2.1 million years ago in what is now Yellowstone National Park. In contrast, predicting even the near future eruption of a volcano such as Mt. Vesuvius is much more difficult. There are too many causally relevant conditions (known and unknown) in the absence of which an eruption won't occur.

As I discussed (Cleland [2002]), the physical source of the asymmetry of overdetermination is controversial. Examples such as an explosive volcanic eruption are commonly attributed to the second law of thermodynamics. The natural processes that produce volcanoes are irreversible; one never sees a volcano literally swallow up the debris it produced and return the land around it to the condition it was in before the eruption occurred. The asymmetry of overdetermination also encompasses wave phenomena, which do not obviously admit of a thermodynamic explanation. Although traditionally associated with electromagnetic radiation (light, radio waves, etc.), the "radiative asymmetry" (as it is known) characterizes all wave-producing phenomena, including disturbances in water and air. It originates in the fact that waves (whether water, sound, light, etc.) invariably spread outwards, as opposed to inwards, as time progresses, which means that the effects of a cause become increasingly widespread in space. Between the second law of thermodynamics and the radiative asymmetry, all physical phenomena above the quantum level (particle and wave) are subject to the asymmetry of overdetermination. While it is tempting to suppose that they are somehow connected—that one is derived from the other, or they are both derived from some third feature of the universe, e.g., its initial conditions at the time of the big bang (Horwich [1987])—the important point, for our purposes, is that they represent objective and pervasive physical features of our universe. It follows that Turner ([2004]) is misguided in claiming that the asymmetry of overdetermination is "strictly metaphysical" (p. 210); it is well grounded in physical theory.

The asymmetry of overdetermination provides the needed objective grounding for the principle of the common cause. According to the thesis of the asymmetry of overdetermination, the vast majority of causal forks open in the direction from past to future. As a consequence the present is filled with epistemically overdetermining traces of past events. This means that it is likely (but not certain) that a puzzling association (correlation and/or similarity) among present-day phenomena is due to a *last* common cause.[10] If the temporal structure of causal relations in our universe were different—if most puzzling associations among events were chance occurrences, or most causal forks opened in the opposite direction (from future to past), or most cause and effect relations were linear (one-to-one) instead of fork-like—one would not be justified in inferring the likelihood of a common cause from a seemingly improbable association among traces. The quest for a smoking gun is a search for additional evidential traces for distinguishing which of several rival hypotheses provides the best explanation for the available body of traces. The overdetermination of the past by the localized present, a physical fact about our universe, ensures that such traces are likely to exist if the initial collection of traces shares a last common cause. For insofar as past events typically leave numerous and diverse effects, only a small fraction of which is required to identify them, the contemporary

[10] The "big bang" of cosmology is of course the earliest common cause of every contemporary phenomenon in our universe, but many subcollections of these phenomena share more recent common causes, including a last common cause.

environment is likely to contain many, as yet undiscovered, smoking guns for discriminating among rival common cause hypotheses.

The asymmetry of overdetermination does not guarantee that every mysterious association among traces is due to a last common cause. As Elliot Sober ([1988], [2001]) and Avi Tucker ([2004]) observe, separate causal processes operating independently may also produce them. Sober cites evolutionary biology as a good source of examples. Bats, birds, and insects, for instance, resemble each other in having wings but do not share a common ancestor with wings; they evolved wings separately. In contrast, lions, whales, elephants, and human females, which have mammary glands, do share a common ancestor with mammary glands. Similarities of the former kind, which are not inherited form a common ancestor, are known in biology as homoplasies, whereas those of the latter kind, which are inherited from a common ancestor, are known as homologies.

But as I have argued ([2008], [forthcoming]) the default preference of historical scientists is for common cause explanation unless they have special purpose theoretical or empirical reasons for thinking that a puzzling association among traces is due to separate causes. I suspect that Sober's contention that it is a mistake for historical scientists to favour common cause hypotheses over separate causes hypotheses (all things being equal) stems in part from his focus on biological examples. Lying in the background of all biological reasoning is Darwin's theory of evolution by natural selection. According to Darwin's theory, similar environments may produce similar adaptations in organisms that do not share a common ancestor with the trait concerned. Furthermore biologists are familiar with numerous examples (e.g., bats and birds) in which this has occurred. It follows that homoplasies pose a very real threat to phylogenetic inferences.[11]

The situation in earth history, however, is quite different. No overarching general theory of geology or planetary science suggests that geological analogies are so widespread as to pose serious threats to common cause explanations for seemingly improbable associations among traces. Nevertheless, a search for a smoking gun for a common cause may turn up empirical evidence that a body of traces was produced by separate causes. A good illustration is radiometric and fossil evidence that the great Permian extinction consisted of two distinct extinction events separated by about 10 million years (Erwin [2006], p. 7). But even in this case the search is ultimately for common causes. Having subdivided the original body of traces into those pertaining to the first extinction pulse and those pertaining to the second extinction pulse, scientists now seek their separate common causes. They also explore the possibility that the two pulses might be causally related through an earlier common cause, e.g., the formation of the supercontinent Pangaea. The point is in the absence of special theoretical or empirical reasons for believing that a puzzling association among traces was produced by separate causes, historical scientists have a good reason for opting for common cause explanations: the asymmetry of overdetermination.

The asymmetry of overdetermination does not guarantee that every past event can be identified from contemporary traces. It is unlikely but nonetheless possible for an event to leave no traces; prime candidates are events occurring before the big bang of cosmology. More significantly with the passage of time the causal information carried by traces becomes increasingly degraded, and eventually may disappear altogether. It is for this reason that a significant portion of historical research is devoted to analyzing and sharpening attenuated traces so that they can be identified and properly interpreted; this often requires the development of sensitive new technologies.

Derek Turner ([2004], [2007]) believes that the threat of "information destroying processes" is so extensive, however, that no interesting epistemological conclusions of the sort that I draw follow from the asymmetry of overdetermination. But Turner conflates information-degrading processes with

---

[11] As I discuss in (Cleland, [forthcoming]), Sober's examples—see his ([2001])—of purely numerical correlations among monotonically increasing quantities (such as British bread prices and Venetian sea levels, which have both been rising over the past two centuries) can be similarly understood. It is true that such correlations are common, but they are also understood to be the product of separate causes (*in virtue of* being purely numerical). It is only when they are treated ambiguously, as being purely numerical but potentially not purely numerical (in which case they are likely to be the product of a common cause), that Sober's admonishment that one ought to suspend judgment between common cause explanations and separate causes explanations when faced with coincidences becomes compelling; numerical correlations among quantities are not puzzling to scientists if they know they are *purely* numerical.

information-destroying processes. While it is true that information becomes degraded over time there is good reason to believe that much of it can nonetheless be recovered with the right technologies. Ancient meteorite craters, for instance, become slowly buried over time until they are no longer detectable from surface features. The Chixulub crater, thought to be ground zero for the impact responsible for the K-T extinctions, was identified by means of aerial surveys of the northern coast of the Yucatan Peninsula utilizing sophisticated geophysical instruments, which revealed a gigantic (at least 170 km in diameter), circular, gravity anomaly buried a kilometer beneath younger sedimentary rock. Analogously, speculation that life on Earth goes back 3.8 billion years rests upon laboratory analyses of carbon isotope ratios in grains of rock as small as 10 μm across weighing only $20x10^{-15}$ g (Mojzsis et al. [1996], p. 56). Remarkably, these analyses reveal an enrichment of the lighter isotope of carbon, which is preferred by life, over the heavier isotope, a correlation that is difficult to explain in terms of non-living processes. Who would have thought that convincing evidence for long-dead microscopic forms of life could be extracted from rocks this old? As these examples illustrate, our ability to extract information about the past from contemporary phenomena is rapidly increasing, so much so that I suspect the twenty first century may become the age of historical science!

Turner nevertheless insists that such cases are the exception rather than the rule. He cites speculation about the colors of the dinosaurs as an example of something that paleontologists will never be able to discover (Turner [2004], pp. 217-8, and [2007], Ch. 2). Admittedly we currently don't know how to determine the color of a dinosaur from its fossil remains. But this doesn't mean that the information isn't there. Indeed, this example bears an uncanny resemblance to the claim that one cannot sex a dinosaur from its fossilized remains. A few years ago this claim would have been just about as plausible but, as Schweitzer demonstrated for a certain T. rex, it is false. In other words, even though information-degrading processes are common, there is little evidence that they completely destroy information as opposed to merely make it difficult to extract. The extent to which information-degrading processes remove identifying information about long-past causes from their traces with the passage of time is an empirical question. If recent technological advances provide any guide, it may be much less than Turner believes. And this brings us to a stunning recent paleontological discovery. While examining a fossilized bird feather under an electron microscope Jakob Vinther and colleagues (2008) recently stumbled upon preserved melanin granules; melanin is a natural pigment that gives color to bird feathers as well as to human skin and hair. The feather was from the Cretaceous, the last age of the dinosaurs. Because fossilized dinosaur skin has been discovered with feathers, the team speculates that they may eventually be able to interpret the color of some dinosaurs from their fossilized remains. This remarkable discovery underscores my central point: the overdetermination of causes by their effects is extensive and pervasive in our universe, and this means that historical scientists can never rule out the possibility of discovering a smoking gun for any hypothesis about the past, however far fetched this possibility may currently seem.

References

Alexopoulos, J. S., Grieve, A. F. et al. [1988]: 'Microscopic Lamellar Deformation Features in Quartz: Discriminative Characteristics of Shock-Generated Varieties', *Geology*, 16, pp. 796-799.
Benner, S. A. and Switzer, C. Y. [1999]: 'Chance and necessity in biomolecular chemistry: Is life as we know it universal?', in H. Frauenfelder, J. Deisenhofer and P. G. Wolynes (eds), 1999, *Simplicity and complexity in proteins and nucleic acids,* Berlin: Dahlem University Press, pp. .
Bohor, B., Foord, E. E. et al. [1984]: 'Mineralogic Evidence for an Impact Event at the Cretaceous-Tertiary Boundary', *Science* 224, pp. 867-869.
Cartwright, N. [1983]: How the Laws of Physics Lie. Oxford: Oxford University Press.
Cleland, C. E. [forthcoming]: 'Prediction and Explanation in Historical Natural Science'.
Cleland, C. E. [2008]: 'Philosophical Issues in Natural History and Its Historiography', in A. Tucker (ed.) *A Companion to the Philosophy of History and Historiography, Oxford*: Wiley-Blackwell, pp. 44-62.
Cleland, C. E. [2002]: 'Methodological and Epistemic Differences between Historical Science and Experimental Science', *Philosophy of Science*, 69, pp. 474-496.
Cleland, C. E. [2001]: 'Historical science, experimental science, and the scientific method', *Geology*, *29*, pp. 987-990.
Clemens, W. A., Archibald, J. et al. [1981]: 'Out with a Whimper Not a Bang', *Paleobiology,* 7, pp. 293-298.
Ereshefsky, M. [1992]: 'The Historical Nature of Evolutionary Theory', in M. H. Nitecki and D. V. Nitecki (eds), 1992, *History and Evolution*, New York: State University of New York Press, pp. 81-99.
Erwin, D. H. [2006]: *Extinction: How life on earth nearly ended 250 mya*, Princeton: Princeton University Press.
Hull, D. L. [1992]: 'The Particular-Circumstance Model of Scientific Explanation', in M. H. Nitecki and D. V. Nitecki (eds), 1992, *History and Evolution*, New York: State University of New York Press, pp. 69-80.
Hempel, C. G. [1965]: *Aspects of Scientific Explanation*, New York: Free Press.
Hempel, C. and Oppenheim, P. [1948]: Studies in the Logic of Explanation, *Philosophy of Science*, 15, pp. 567-579.
Horwich, P. [1989]: *Asymmetries in Time*, Cambridge, Mass: MIT Press.
Jeffares, B. [2008]: 'Testing Times: Regularities in the Historical Sciences', *Studies in the History and Philosophy of Biological and Biomedical Sciences,* pp. 469-475.
Johnson, K. R. and Hickey, L. J. [1990]: 'Patterns of Megafloral Change Across the Cretaceous-Tertiary Boundary in the Northern Great Plains and Rocky Mountains', in V.L. Sharpton and P. D. Ward (eds), *Global Catastrophes in Earth History*, Boulder, CO: Geological Society of America, Special Paper 247, pp. 433-444.
Kerr, R. A. [1987]: 'Asteroid impact gets more support', Science, 236, pp. 666-668.
Kleinhans, M. G., Buskes, C. J. J. and de Regt, H. W. [2005]: '*Terra Incognita*: Explanation and Reduction in Earth Science', *International Studies in the Philosophy of Science*, 19, pp. 289-317.
Leather, J., Philip, A. A., Brasier, M. D., and Cozzi, A. [2002]: Neoproterozoic snowball Earth under scrutiny: Evidence from the Fiq glaciation of Oman', *Geology*, 30, pp. 891-894.
Lewis, D. [1991]: 'Counterfactual Dependence and Time's Arrow', *Nous*, 13, pp. 455-476.
Mitchell, S. D. [2002]: 'Ceterus Paribus—An Inadequate Representation for Biological Contingency', Erkenntnis*,* 57, pp. 329-350.
Mitchell, S. D. [2000]: 'Dimensions of Scientific Law,' *Philosophy of Science,* 67, pp. 242-265.
Mojzsis, S., Arrhenius, G., McKeegan, K. D., Harrison, T. M., Nutman, A. P., and Friend, C. R. L. [1996]: 'Evidence for Life on Earth before 3,800 Million Years Ago', *Nature*, 385, pp. 55-59.
Officer, C. B. and Drake, C. L. [1985]: 'Terminal Cretaceous Environmental Events', *Science,* 227, pp. 1161-1167.
Powell, J. L. [1998]: *Night Comes to the Cretaceous*, San Diego: Harcourt Brace & Co.
Reichenbach, H. [1956]: *The Direction of Time*, Berkeley: University of California Press.

Ruse, M. [1971]: 'Narrative Explanation and the theory of Evolution', *Canadian Journal of Philosophy,* 1, pp. 59-74.
Sober, E. [2001]: 'Venetian Sea Levels, British Bread Prices and the Principle of the Common Cause', *British Journal of Philosophy of Science*, 52, pp. 233-50.
Sober, E. [1988]: *Reconstructing the Past*, Cambridge, Mass: MIT Press.
Schmitz, B. and Asaro, F. [1996]: 'Iridium Geochemistry of Ash Layers from Eocene Rifting', Bulletin *of the Geological Society of America,* 108, pp. 489-504.
Schweitzer, M. H., Wittmeyer, J. L., and Horner, J. R. [2005]: 'Gender-Specific Reproductive Tissue in Ratites and *Tyrannosaurus rex*. *Science'*, 308, pp. 1456-1459.
Tucker, A. [2004]: *Our Knowledge of the Past*, Cambridge: Cambridge University Press.
Turner, A. [2007]: *Making Prehistory*. Cambridge: Cambridge University Press.
Turner, D. [2004]: 'Local Underdetermination in Historical Science', *Philosophy of Science*, 72, pp. 209-230.
Vinther, J., Briggs, D. E. G., Prum, R. O. and Saranathan, V. [2008]: 'The color of fossil feathers', *Biology Letters*, 4, pp. 522-525.
Ward, P. D. [1990]: 'The Cretaceous/Tertiary Extinctions in the Marine Realm; a 1990 Perspective', in V. L. Sharpton and P.D. Ward (eds), 1990, *Global Catastrophes in Earth History*, Boulder, CO: Geological Society of America, Special Paper 247, pp. 425-432.
Ward, P. D. [1983]: 'The Extinction of the Ammonites', *Scientific American*, 249, pp. 136-147.

Hayek and Popper on the Evolution of Rules and Mind

Gerald J. Postema[♣]

Questions about the nature of informal social rules have become a major focus of attention in legal philosophy and legal theory in recent decades. Since Hart insisted that modern municipal law rests on a fundamental *social* rule, the rule of recognition practiced by law-applying officials, it has been a preoccupation of much of analytic legal philosophy to explain the nature of social rules (or "conventions"). More recently, theorists have come to recognize that an understanding of international law cannot proceed very far without a solid understanding of informal rules ("custom"), for customary law still plays an important and foundational role in that domain. Also, in legal theory there has emerged recently a multi-disciplinary study of what are called "social norms," which arguably fall into another species, along with "customs" and "conventions," of the genus "informal social rules." Similar interest in informal social rules can be found in contemporary moral and political philosophy, often drawing conceptual resources and explanatory frameworks from game theory and socio-biology. In these fields, attention has especially turned to explaining how informal social rules emerge and change. An account of the origin and dynamics of such rules is thought to be fundamental to our understanding of how they function, which, in turn, informs our understanding of law, morality, and political institutions. This focus of philosophical attention is not new, of course. It finds at least one especially perceptive antecedent in the eighteenth century in Hume's attempt to explain the "origins" of justice and allegiance in his *Treatise of Human Nature* [Hume 1998a].

Karl Popper did not directly address this set of theoretical issues, but his work offers some surprising insights into the evolution of human rational capacities that may be of use to those seeking an account of the dynamics of social rules that is open to the full range of resources that human beings may bring to bear in shaping rules for their lives together. Friedrich Hayek, on the other hand, developed, in a number of works, a systematic explanation of the emergence and dynamics of informal social rules. His theory repays study by those who are interested in contemporary discussions of the evolution of social rules, while its weaknesses invite consideration of Popper's view of the domain of "objective knowledge" and his account of the evolution of mind, or so I shall argue here. Hence, I propose to look at the views of Hayek and Popper on the evolution of mind and social rules, in the hope that, taking them together, we can gain some insight into issues that have been at the center of attention in much recent moral, legal, and political philosophy.

I begin with an exposition of Hayek's framework for explaining the dynamics of what he calls "grown order," followed by a discussion of problems that threaten to undermine his explanatory scheme. I end with a consideration of ideas suggested by Popper's work that, although not entirely welcome to Hayek, might put us in a better position to solve the problems that prove intractable on Hayek's theory.

The Dynamics of Grown Order: Hayek's Explanation of Social Rules and Institutions

Central to Hayek's social and legal theory is his explanation of the nature and dynamics of social rules and institutions, especially the market and law, as what he calls *grown orders*. His explanation is multi-layered, accounting for the social rules and institutions we see in terms of deeper levels which are increasingly less obvious to the casual observer. To understand the typical operation and dynamics of social rules, he argued, we must look more closely at the operation and dynamics of rules of the mind. The superstructure, as it were, of social institutions builds on, but never transcends, a rich and layered substructure of more basic rules of conduct and of thought. Our complex social institutions are anchored in this more basic substructure and are radically dependent on it.

[♣] Cary C. Boshamer Professor of Philosophy and Professor of Law, University of North Carolina at Chapel Hill

Hayek's systematic explanation relies on three fundamental ideas: the idea of a rule (and rule-following), and the "twin ideas" of spontaneous order and evolution, the two components of his idea of grown order. The three ideas are interdependent parts of a single, integrated explanatory scheme, designed to show that key elements of social life are ordered—not the product of some designer, but rather the "unintended consequences" of impersonal and external forces operating on behavior and thought of human beings directed to other ends and purposes. I begin with Hayek's notion of rules, because the other two notions work with an idea of social order regarded as the product of behavior directed by rules.

Hayek on Rules

Rules as subject-grasped and subject-directing patterns

For his purposes, Hayek deploys a very broad concept of a rule. Rules, as he proposes to use the concept, direct both thought and conduct. Rules of thought or mind concern matters of immediate perception, as well as judgment and higher order structures like mathematical concepts and abstract theories [Hayek 1967, 23-24, 43-46]. "Perception," for Hayek, includes everything from immediate sensory judgments (e.g., a rhythmic pattern of lights going on and off) to organized, albeit very particular, judgments about one's situation (e.g., the entrepreneur's sense that a certain product might succeed in the current market). Rules falling along this wide spectrum, on Hayek's proposed understanding, are (i) recognized or *projected patterns*—configurations of items or elements that are grasped (at least in the most basic forms) as a *Gestalt*—that (ii) generate a determinate *response* in the subject [Hayek 1967, 23, 45, 52; Hayek 1973, 75]. This proposal needs unpacking. Four elements of this proposal call for our attention.

First, rule-patterns exist only as "recognized" [Hayek 1967, 23, 45], although this "recognition" need not be conscious, let alone articulated by the subject. That is, the patterns may be grasped by the mind of the subject—they are "in" or "of" the mind—at a pre-conscious level, entirely "without intellection," on the one hand, or as matter of conscious mental construction, on the other. Thus, the patterns are not, strictly speaking, detected in the nature we perceive, but rather are responses to encounters with nature and projected onto it. Second, rule-patterns are always *abstractions*, patterns resulting from selecting and ordering certain elements and ignoring others of experience [Hayek 1973, 30]. Moreover, rule-patterns are, in Hayek's view, *generic* in the sense of being logically universal, but also in the sense of comprising many real (not merely logically possible) circumstances and instances.

Third, the grasped patterns are always accompanied by *dispositions* of *response*, either a disposition to *see*, *feel,* or possibly to *judge* something, or a disposition to act in a certain patterned way [Hayek 1973, 75, 79]. It is not entirely clear whether Hayek's view is that the pattern causally generates the disposition or that what the subject experiences is a *patterned disposition* to perceive or to act. I suspect that he thinks that, at least at the most basic levels, the latter is true, although at higher levels there may be room to distinguish the pattern grasped and the subsequent judgment or action. In any case, Hayek's rules are not inert: they are *determinants* of thoughts and especially behavior [Hayek 1973, 79]. Rules are not merely grasped by the subject, they *direct* the subject.

This supplies the foundation for Hayek's account of rule-following. At its most fundamental level, for Hayek, rule-following is a matter of "know how,"[12] that is, a disposition to act (perceive, judge) in a certain way, arising from a situation that, having grasped its significance, disposes one so to act. Rules, on this view, are never merely regularities or patterns; rather, they are *grasped* patterns that are or give rise to dispositions.[13] Thus, for Hayek, rules are *subject-grasped* and *subject-directing* patterns. Moreover, when we observe rule-generated behavior of things in our environment, including the actions of other agents, we understand each instance of the grasped regularity as having a common

[12] Hayek draws heavily on Ryle's [1949] distinction between "knowing how" and "knowing that."

[13] In his essay, "Rules, Perception and Intelligibility," Hayek traces our rule-following capacity to our nervous system that acts as both "a movement pattern *detector*," recognizing actions conforming to rules, and "a movement pattern *effector,*" generating those actions in appropriate circumstances, or at least disposing us to act in those ways [Hayek 1967, 45].

cause; not merely presenting us with a pattern, but also manifesting a rule-governed order. Rules manifest themselves in such regularities.

Fourth, since rules of conduct are dispositions to act, rather than actual patterns of actions, it is possible for a subject to be directed to act by a rule and yet not act on the rule. This will occur when the conditions for the realization of the disposition are not met; and, in Hayek's view, among the most important conditions for realization of a rule-disposition is the condition that there is not some other disposition operative in the subject which prevails at the time of action. Not only do rule-dispositions live in subjective environments along-side other rule-dispositions, but, in Hayek's view, rules of thought and conduct always exist together in complex networks. Rules are able to do their work, with subtlety and flexibility, in part because they get their content and meaning from their place in a system of rules, forming "chains" of interconnected meanings [Hayek 1967, 57-8]. This is as true of our patterned responses at a very primitive psychological level as it is of sophisticated patterns of reasoning and codes of law.

The Primitive Evolution of Rules of the Mind

Rules of thought and action, in Hayek's view, *constitute* the mind of human individuals [Hayek 1973, 18, 30]. They are products of the encounters of human individuals with their natural and social environments [Hayek 1973, 17-18]. Hayek's account of the basic process by which "rules of the mind" are formed is an early form of what is now called "connectionism" or "neural network" theory [Gaus 2006, 248-52]. Roughly, the view is that, at the most primitive level of formation, rules of the mind are the effects of external causes on an individual's sensory apparatus. The external world causes certain neural responses with some determinate configuration. The mind "grasps" a pattern when two events trigger the same configuration and that configuration, further, yields some response on the part of the individual, either a phenomenal experience or behavior. So, for an individual to learn a rule or pattern is just for there to be established in that individual's brain a neural pathway that is triggered by multiple external events.[14] These external, connection-establishing events have their sources in the physical environment or the social environment of individuals. We can expect different individuals to have relevantly similar experiences and responses to the external world to the extent that they interact with a similar environment (and the causal mechanism by which neural networks are established works in a similar manner in those individuals).

This potential for overlapping rules of the mind is reinforced by the influence of the social environment on the development of the mind. The root of social "learning" is the inborn capacity of human beings (and other higher animals) to mimic the behavior of those around them [Hayek 1967, 47-48]. Following the lead of eighteenth-century Scots,[15] Hayek observes that, even very young infants, without the benefit of a mirror to observe their own movements, are able to reproduce the movements or gestures of those around them. These mimicked movements establish a network which is triggered later by other behavior registered by the individual as similar.

This provides the basis for learning of routines of action and of perception and thought. And this learning not only crosses sensory modalities (as in the primitive case across sight and kinesthetic modalities), but also boundaries between persons. Rules are "grasped" simply by being enacted, as it were, in the behavior of the individual learning them, where "enacting" means that a disposition to respond is established by the neural configuration established.[16] Moreover, Hayek observes, in our

[14] Note that on this view we cannot infer that the external world comes already patterned, but only that the mind responds to stimuli by organizing them into patterns. "Abstraction," Hayek insists, is not an advanced activity of the mind but rather absolutely the most basic and primitive. The mind's initial response to the external world is a pattern-forming response [Hayek 1973, 30].

[15] Hayek cites Dugald Stewart and Adam Smith, but the locus classicus of this line of psychological observation is Book II of Hume's *Treatise of Human Nature* [Hume 1998a].

[16] Hayek does not explicitly acknowledge the further important fact that it is typical of human learning that the "similarity" of the responses is not simply a matter of the events triggering the same neural configuration, but of its being recognized by others and that recognition being recognized by the learner. In this way, a distinctively social form of learning, not

social environments we not only learn certain common behavioral routines, but also their meaning. We perceive in the movements of others their mood or attitude that makes the movements intelligible to us [Hayek 1967, 55, 59]; we grasp the behavior as "purposive" rather than random, not just behavior with some regularity, but rule-following behavior. This in turn enables a degree of understanding across minds, a basis for a degree of *Verstehen* [Hayek 1967, 58-60].

The Implicit Dimensions of Rules

A core feature of Hayek's theory of rules is his doctrine, repeated with the frequency of a mantra, that subjects directed by rules of thought and action need not be, and indeed predominantly are not, aware of these rules. The rules are implicit, matters of only tacit understanding. They are, as he says, "known by none, but understood by all" [Hayek 1967, 46]. Hayek's doctrine of the implicit dimension of rules comprises several related claims, some based on observation or argument, some asserted but never adequately defended.

He begins from the observation that we are able to act on very sophisticated rules without even the slightest awareness of them. His favorite example is that of children who manage to use language with great facility without any awareness of its rules of diction, grammar and syntax, let alone a capacity to articulate those rules [Hayek 1967, 42-44]. He then goes on to maintain repeatedly that this is true about the vast bulk of our (patterned, rule-governed) knowledge of our physical and social world. Moreover, not only are these rules (hence, this knowledge) currently unarticulated, but the vast bulk of it cannot be articulated or even brought to our awareness. This is, in part, due to the fact, as Hayek sees it, that most of these rules are highly localized, restricted to specific times, places, and circumstances of individuals and embedded in the particular activities and skills of their ordinary practical lives. These rules are so deeply embedded in their practice that they cannot be brought to consciousness without distilling away most of their content. The problem, it seems, lies in part in the fact that something can be made explicit to consciousness, in Hayek's view, only if it is articulated linguistically, and we lack the resources to articulate the content linguistically. But it is due even more to the fact that we could not capture it even if our linguistic resources were far more sophisticated, because it is so vast, interconnected, and embedded in practice. Thus, inevitably, a very large part of the whole which gives determinate meaning to any given rule in particular circumstances remains inaccessible to the agent who learns how to follow it. Moreover, since we are unable to make this knowledge explicit, Hayek concludes, we also cannot share it. It is widely dispersed and in very large measure private.

This conclusion rests on a very strong assumption of *subjectivism* (maintained despite the potential for a more modest version represented by his recognition of the possibility of *Verstehen*). This very general and deep assumption takes various forms in Hayek's work. For our purposes it surfaces in two forms. (a) In his epistemology, subjectivism takes the form of the claim that "knowledge exists only as knowledge of [i.e., possessed by] individuals" [Hayek 1960, 24]. He rejects any idea of social, shared, or common knowledge.[17] (b) The second form in which subjectivism

(Contd.)

matched by learning in a subject's physical environment, takes place. This, in Hume's view, is one key source of the human capacity for what he calls "sympathy".

[17] Although he rejects any notion of shared or common knowledge, he does insist both that widely dispersed and largely private knowledge is nevertheless indirectly *available* to individuals and that this knowledge is *embedded* in rules. The market is Hayek's favorite example of how dispersed and private knowledge is nevertheless socially available. The market is a framework of rules that serves to coordinate the actions and interactions of countless numbers of agents, each market player acting on his or her own local knowledge. The price system does not itself "contain" within its mechanism the knowledge of each player, but it enables each to adjust their decisions to the knowledge abstractly represented by the prices offered. Prices are, as it were, content-independent markers of dispersed knowledge that remains essentially unarticulated and inaccessible in any more direct form. In market economies, knowledge drives activities of parties without the need for any central accumulation of that knowledge. Such knowledge remains dispersed. It is never common.

Similarly, Hayek maintains that knowledge is "embedded" in social rules just in the sense that the process by which they have evolved ensures that they are tested against a wide range of circumstances and have proved to be adequate adaptations (adequate for group effectiveness, as we shall see) to those circumstances. Again, there is, strictly speaking, no accumulated wisdom of the ages stored in these rules; rather, the rules are simply the product of an impersonal (and

surfaces is in his understanding of methodology of social explanation.[18] In Hayek's view, our understanding of social rules and institutions must not only start from, but must also always be ultimately reducible to (or brought back home, in some other suitably strong sense) to statements about the mental states of individual subjects. Hayek recognizes that it is possible for us to gain some understanding (*Verstehen*) of each other, but this involves grasping what another mind desires, intends, or "means" [Hayek 1967, 58-60]. It is a matter of grasping something in or of the mental state of another person and this grasp will always be subject to very severe limits, because we can only grasp what is conscious to the other person and that, as we have seen, is only a very small portion of the basis of their perceptions, judgments, desires, and purposes.

Two Explanatory Models: Spontaneous Order and Evolution

Commentators and critics often treat Hayek's evolutionary explanation of social rules and order as independent of and in competition with explanations drawn from the idea of spontaneous order. This, I think, is a mistake. In Hayek's eyes they are interdependent explanatory schemes.[19] His account of the evolution of social rules depends heavily on the idea of spontaneous order and the role of rules in producing that order; moreover, the sources of disequilibrium, and hence innovation (mutation) and reproduction (replication), needed for the evolutionary story, occur within the process explained by the spontaneous order scheme. At the same time, the notion of spontaneous order presupposes rules that direct the behavior of individuals from which the order emerges, rules that, on Hayek's account, have emerged from an evolutionary process; moreover, spontaneous order explanations are incomplete explanations of the dynamics of social order and social rules, because (except within certain limits) spontaneous order explanations are static, such that when internal forces no longer suffice to bring the disorder back into disequilibrium spontaneous order explanations run out and the evolutionary account must be deployed to explain how a new order is established. Thus, to understand Hayek's proposal for explaining social institutions, we must relate these two explanatory schemes. Although they are interdependent, it does not distort them too much to view them as working in two stages, following the trajectory of a spiral rather than a vicious circle. The image of a spiral also allows us to capture the idea that the interdependent processes of spontaneous order and evolution build on previous and in some cases more basic stages.

Spontaneous Order Explanations

We are tempted to regard all manifestations of order in nature and social life as products of design-governed efforts; however, Hayek argues, many ordered structures must be explained as undesigned, endogenous, and self-generating. The "order" or observed pattern emerges from the interaction of a large number of elements responding to their environment (including the behavior of other elements) according to certain forces or rules that direct that behavior. Hayek's model of spontaneous order applies to both natural phenomena like the formation of crystals or patterns of iron filings and social phenomena like a living language or the market [Hayek 1967, 39-40; Sugden 1998b, 485]. I will focus here on spontaneous social orders.

To begin, Hayek distinguishes between rules of conduct and the social order that they (indirectly) generate [Hayek 1967, 66-69]. "Order" as Hayek understands it, is that "state of affairs in which a multiplicity of elements of various kinds are so related . . . that we may learn from our acquaintance with some . . . part of the whole to form correct expectations with regard to the rest"

(Contd.)

for that matter, content-independent) process which nevertheless offers promise of a substantial degree of success in day-to-day social interactions.

[18] A classic statement of Hayek's methodological subjectivism can be found in *The Counter-Revolution of Science*: "Not only man's actions towards external objects, but also all the relations between men and all social institutions can be understood only in terms of what men think about them. Society as we know it is, as it were, built up from the concepts and ideas held by the people; and social phenomena can be recognized by us and have meaning to us only as they are reflected in the minds of men." [Hayek 1952, 34-5.]

[19] This interdependence is clearly evident in Hayek 1967, ch. 4. Heath (1992) and Gaus (2006) are rare among readers of Hayek to recognize this interdependence and only Gaus, in my view, comes close to understanding the nature of this interdependence.

[Hayek 1973, 36]. Social order is an emergent property of the actions and interactions of a large number of agents, that is, it is an abstract pattern manifest in the interactions of particular individuals which may persist even if all the individuals are replaced by others [Gaus 2006, 233-4]. This pattern is the product of (i) the actions of large numbers of individuals (ii) in an environment of a determinate nature which (iii) includes the actions of others, all of whom (iv) respond to local knowledge of that environment (v) from a potentially wide variety of motives (vi) within the limits defined by the system of rules in force in the group. This order is "spontaneous" because it is the result of individuals arranging themselves according to "forces" (in the social context: motives within the framework defined by rules) in a specific environment. The order is the resultant of the balance of these forces [Sugden 1998, 487]. Because of the interaction among the individuals and the feedback from this interaction, the properties of the order are not simply the aggregate of the properties of individual elements, but rather are emergent from them.

The relationship between the rules operative in a given social context and the order that emerges is indirect and complex, because the order emerges from the combined influence of the rules and the environment on the choices and consequent interactions of the agents. Thus, it is not the case that every set of rules can be expected to produce a corresponding order; indeed, some rules may prevent any order from forming or may produce an order that is dysfunctional from the point of view of the group or of its individual members [Hayek 1973, 43]. Also, it is possible that the same social order may be produced by different sets of rules and that the same set of rules may yield different social orders [Hayek 1967, 67-68; Hayek 1973, 43-44]. A change in the rules may not result in a change in the social order and rules may produce very different orders in environments that have significantly different properties. The environments in which individuals interact and the way in which and the extent to which the rules influence the actions of individuals, greatly affect the relationship between rules and the resulting social order or disorder. Finally, a social order may emerge even if the regularity in the behavior of individual members of the group is not universal. This lack of uniformity can be of two broad kinds. (i) The rules may call for quite different routines of conduct from different people in different roles, stations, or circumstances. What is important for social order is not uniformity of behavior across the membership of the group, but its coordination. Coordination requires only that the rules be broadly compatible, making possible coordinated behavior of those directed by them.[20] (ii) There may be some, perhaps even a substantial, degree of irregularity of behavior (i.e., deviations from the rules) in the group. Just how must irregularity or deviance a social order can tolerate is determined by a wide variety of factors, some environmental, some psychological, some having to do with the internal relations among the rules. Moreover, Hayek realizes that there must be some degree of this kind of irregularity if the social order is to have the flexibility needed to cope with exogenous shocks, and to permit endogenous changes, that cause adaptive changes in the order to occur.[21]

Sugden and Gaus have identified several salient features of spontaneous orders as Hayek conceives of them. First, they are *path-dependent* in the sense that the properties of the order at any point in time depend on its history [Sugden 1998b, 488; Gaus 2006, 233]. Second, they *approximate,* but never strictly achieve, *equilibrium* [Gaus 2006 234], with the result that there is always some greater or smaller degree of disequilibrium in the system. Nevertheless, third, spontaneous orders are, within limits, *self-maintaining* [Gaus 2006, 234]; that is, it can survive exogenous and endogenous shocks, restoring its (approximation to) equilibrium. Finally, Hayek recognizes that the spontaneity of a social

[20] This is a consequence of Hayek's view that rules come linked together in integrated packages, rather than merely aggregated in sets. But the idea is in tension with his official view that rules of spontaneous orders are "abstract" both with respect to the ends served by them and with respect to the agents and circumstances to which they apply [Hayek 1967, 56; Hayek 1973, 50].

[21] This flexibility, as we shall see, is essential to his scheme of evolutionary explanation, but again this feature is in tension with his frequent insistence that rules must be followed *rigidly* [Hayek 1967, 90-91]. The latter dogma, like so many other statements of Hayek's, is overly broad and incautiously formulated for his primary purpose, which at the point in the text was to counter the idea that rules are mere rules of thumb for agents who are act-utility maximizers (governed by "expediency" rather than "principle") [Hayek 1973, ch. 3].

order is a *matter of degree*.[22] As Sugden points out, spontaneity is a function of at least two properties: dispersion of power and redundancy [Sugden 1998b, 487]. If we understand "power" to be the extent to which an individual can influence the properties of the social order, then we can see that the more widely power is dispersed over a population, the less power each individual will have; and, thus, the greater the dispersion, the greater will be the spontaneity of the order. Similarly, spontaneity is in part a function of the density (the number and overlapping nature) of the relations among the members of a group and the interchangeabilty of the parts. Together, these yield redundancy in a system: the greater the redundancy in a system the less likely is it to be affected by deviations of small numbers of the members. Dispersion of power and redundancy admit of degrees, and, as a consequence, so will the spontaneity of an order.[23]

With these general features of spontaneous orders in mind we can get a sense of the nature and limits of the dynamic movement within spontaneous orders as Hayek understands them. First, changes in the environment in which members interact (exogenous shocks) may lead to adjustments of behavior of the members within the parameters defined by the existing rules, thereby re-establishing the order. We might call this a case of simple self-maintenance of the order. It is also possible that exogenous shocks, or endogenous challenges to the rules, result in change of the rules. This change may not produce an overall change in the order, in which case we have a more complex form of self-maintenance, or the change of the rules may be substantial and influence the integrity of the social order. Changes of some rules may bring about shifts in other rules of the system and these adjustments may restore the (near) equilibrium of the order. Other changes in the environment or changes of the rules may require substantial adjustments in the behavior of members of the group thereby altering the nature of their interactions. In that case, the emergent social order will also change, resulting over time either in disorder or the emergence of a new order with different properties.

In each of these cases, individual members may be affected, as may the felicity and fortunes of the group as a whole. Hayek makes clear that the fact that an order emerges spontaneously from the interactions of a group does not guarantee that the order is beneficial, let alone optimal, either to individual members or to the group as a whole [Hayek 1967, 67; Hayek 1973, 43-44]. Indeed, it is possible that a set of rules may even prevent order from emerging, or bring about damaging and socially dysfunctional disorder. Hayek's notion of order is entirely value neutral and the fact that an order has arisen spontaneously implies no special value and offers no guarantee of its being in any way beneficial.[24] There is nothing in the idea of spontaneous order to ensure that a coordinated order will be achieved through interactions of individuals (even if they are directed by or act within the limits of rules), neither will it insure that order, once achieved, will be maintained. If social interaction achieves and maintains order, this will be due in part to forces outside those operative within spontaneous orders.

Thus, the idea of spontaneous order provides a model for explaining the emergence and alteration of social rules. On this model, new rules emerge and are altered in response to changing environmental conditions or in response to changes in rules that result from the irregular behavior of some individual members. The balance of forces within the order brings about these changes, without the intervention of any designers who have a view of the whole system of rules and the order it tends to produce. But we are left with several major questions. One question is: where do the rules that initially structure the spontaneous order originate? Another is: what more precisely is the process by which rules are adjusted in response to exogenous and endogenous shocks? How are rules changed? What role does the judgment or practical reasoning of the individual members play in responding to

---

[22] Hayek 1973, 41-42, 45-46. His penchant for sharp dichotomies, especially between made/imposed order (*taxis*) and grown/self-generating order (*kosmos*), often obscures this fact, perhaps not entirely unintentionally.

[23] The important conclusion we must draw, although Hayek obscures it in many ways, is that we may have to ask what degree of spontaneity of a social is desirable, and what reasons ought to guide that choice.

[24] Of course, later Hayek seeks to link spontaneous orders with individual freedom, but that is not part of his initial construction of the idea of spontaneous order as an explanatory device. And this extension of the concept of spontaneous order depends on principles or evaluative premises that are not at the core of the notion itself.

the shocks, or in creating those (endogenous) shocks?[25] And how are rules that actually prevail selected, if, by hypothesis, this is not done by individuals taking account of the impact of changes of the rules on the social order as a whole? What reason have we to think that some order will be achieved and that it will be in some sense beneficial? For answers to these questions Hayek directs our attention to his account of social evolution [Hayek 1973, 44], the necessary complement to the explanatory structure provided by the idea of spontaneous order.

Evolution of Social Orders and Social Rules

Hayek's account of the evolution of social rules and institutions is a generalization of the Darwinian account [Hayek 1967, 32], but the precise components and mechanisms of Hayek's account are difficult to pin down. This is largely because never in his many discussions of social evolution does he offer a careful, systematic statement of his theory and it is difficult to reconcile the many different partial accounts one finds scattered in his work. This is not the place to attempt to reconcile all these passages; rather, I will offer a reconstruction that seeks to remain faithful to the central motivation of Hayek's account in the hope that is results in a plausible version of that account.

Every evolutionary explanatory scheme must provide (i) a mechanism of selection, including specification of (a) the unit of selection and (b) the basis for selection, and (ii) a mechanism of change, including accounts of (a) the source of innovations or variation ("mutations") and (b) their reproduction ("replication") in the population. Let us look at these elements in order, beginning with the mechanism of selection.

Unit of Selection. On Hayek's account, the evolutionary selection of social rules is indirect, the result of evolutionary forces operating on the social order as a whole; that is, the selection social rules is the result of *competition among social orders* [Hayek 1967, 71]. However, this does not yet determine the unit of selection; it does not tell us whether we should look for the effects of changes in the social order on the felicity and functioning of individual members or on that of the group as a whole.[26] Hayek is often criticized for being inconsistent about whether he favors a group or an individual adaptiveness criterion [e.g., Heath 31-33]. Gaus maintains that Hayek embraces both independent accounts and holds that social orders are subject to two competing forces of evolutionary selection [Gaus 2006, 240-46]. I believe, however, that Hayek thought that the two elements are closely integrated in a single explanatory account. Impacts on and competition among groups and among individuals are both important for his unified account and are inseparable.[27] For, on his view, groups have no aims and enjoy no benefits of their own apart from the aggregate good of individual members. At the same time, individuals benefit—they "succeed" in carrying out their ends and aims, as he likes to put it—only when there is an effective social order that coordinates their efforts and interactions with other members of the group. Moreover, it is judgments of relative "success" of individuals acting within the framework of a given system of rules that is an important part of Hayek's account of the mechanism of change in his evolutionary story. We have no guarantee, of course, that a given social order that functions well for the group as a whole will prove optimal or even beneficial for each individual member; so there is room for familiar problems of collective action to complicate the individual/group relationship and critics are quick to point out that the evolutionary process that depends on group selection can be undone by such collective action problems. However, Hayek tends to downplay the potential conflict over the distribution of the benefits of this group success and the possibility of substantial opportunities for individuals to ride free on the cooperation of others. We will explore his reason for doing so in Part II. B.

Putting this issue aside, it is possible to say with reasonable confidence that, although he thought individuals play an important role in the evolutionary process, Hayek took the unit of selection to be the social group [Hayek 1967, 67-68, 71-72; Hayek 1973, 9, 17-19, 74, 99, *passim.*]. The effects

---

[25] That is, how do changes in the group's rules occur and to what extent are these changes the product of deliberate choices and actions of members of the group? These questions lie at the core of both Hayek's account of spontaneous order and of the companion theory of the evolution of social rules.

[26] It also leaves unspecified how the relevant group is to be determined, but I will ignore this indeterminacy.

[27] See Hayek 1973, 18 and 80 where the two are closely linked—or, as some critics would have it, confused.

on individual felicity and functioning play an important role in the process, as we shall see, but evolutionary forces of selection work at group level in Hayek's model.

Basis of Selection. Social orders are selected by evolutionary forces, according to Hayek. The basis for selection is primarily and ultimately the "success" or "effectiveness" (sometimes he says "efficiency") of the group the interactions of whose members tend to manifest a certain order relative to and in competition with the relative success of other groups in the vicinity [Hayek 1967, 67, 71-2; Hayek 1973, 11, 17, 80, 99]. Hayek does not specify this criterion of "success." We are told that the more successful groups "prevail over" or "displace" their competitors [Hayek 1967, 70; Hayek 1973, 9, 18, 99], or perhaps simply grow and, thereby, are better able to produce wealth and conditions for decent life for their members [Hayek 1973, 80]. Some readers contrast group survival with group growth [Heath 32-33; Gaus 2006 240-43], but Hayek seems to think these are closely related. Groups "prevail" over other groups, he maintains, not necessarily through a clash of forces and the literal destruction of rivals, but rather through doing a better job of enabling individuals to achieve their goals. They, thus, tend to attract members of other groups, leading eventually to the demise of the rival or its assimilation into the more successful group [Hayek 1973, 169 n7]. Hayek's basic thought seems to be that it is through doing a better job of guiding expectations and coordinating interactions of individual members than their rivals, groups that grow stronger, wealthier, and more powerful are then able, either through conquest or through assimilation to win in competition with rival groups. Hayek does not rule out evolution that is red in tooth and claw, but he seems to think that it more typically proceeds in a more pacific manner.

Groups prevail in virtue of the properties of their social orders, properties which are products of interaction structured by the groups' rules of conduct. Thus, social rules are selected for their contribution to the "success" of groups (consisting of the "success" of their members). Notice two key features of this account of the selection of rules. First, it moves entirely without design at the social level (although it may involve a vast number of locally oriented, goal-directed decisions and choices by individuals). Selection operates at the group level, but no agent or collectivity of agents decides or acts at that level. Second, rule selection is relative to several conditions: (i) to the system of rules of which it is a part, and hence to the history of the development of those rules, (ii) to the environmental conditions in which the group which practices the rules must function, and (iii) to the groups that happen to be in the vicinity at the time and in the environment and that compete with that group. Thus, there is no basis for concluding from the stable existence of a system of rules in a group at a given time that those rules are optimal, or optimal for that group, or even for that group in that environment. For the only rules tested are those that in fact developed historically in the prevailing group and in its rivals. We cannot even conclude for evolutionary success that the rules operative in a group at a given time are superior to those of its past. Past rules might actually be better for the group, but they may no longer be available, given the evolutionary history of the group [Gaus 2007, 163-4].[28]

Mechanism of Change—the process of innovation and replication. Evolution is a dynamic process, so in addition to the mechanism for selection, we need an account of the forces that introduce and replicate changes which then may bring about changes in the social order on which the forces of selection operate. We can gain a sense of what is needed at this point in the explanatory structure by looking at the analog of species evolution. Biological evolutionary forces operate on traits of individuals of a given genotype and changes of the genotype result from mutations of genes in individuals, which are then passed on to other individuals through reproduction. Changes producing individual traits that enable the species better to meet the challenges of its environment are selected; those that do not, die with the individuals and their offspring that carry the mutations. Evolutionary adaptations of a species depend on just enough flexibility of the genetic structure to allow for mutations combined with sufficient rigidity to insure that mutations are transmitted with fidelity to other individuals of the species. The mutations come from random, exogenous influences on the genes as new individuals are produced.

[28] Thus, again, if Hayek wishes to draw conclusions about the rationality or merits of evolved rules, he must do so on the basis of premises not included in this explanatory scheme.

In Hayek's model of social evolution social order is the analog of individual traits and rules play the role of genes. Thus, Hayek's model needs (i) rules with some degree of flexibility, (ii) a process by which variations in rules can arise, and (iii) a process by which variations are transmitted to other individuals in sufficiently large numbers that they can have some impact on the social order as a whole. Hayek has something to offer on each of these points.

First, Hayek maintains that social rules, although they call for strict adherence, are "voluntary" in the sense that deviations are possible and are not so severely sanctioned that individuals never have an incentive to consider deviation [Hayek 1960, 63]. Or perhaps we should say that social rules are adaptable to the extent that they enjoy this flexibility. Second, changes in the rules, on Hayek's model, are the results of decisions and actions of individuals seeking to realize their goals with a view only to local circumstances and local effects of their actions within the framework of the established rules. Changes arise from individuals engaging in "trial and error" testing, which can have its roots either in *mistakes* or intentional *experimentation*. Actions that deviate from the rules are assessed in terms of their relative success in furthering the goals of the agent. Deviations have their causes in changes in the environment or in individual's imagining new ways of adjusting to the existing environment. While these changes may influence the social order as a whole, individuals respond only to local conditions without appreciation for such systemic effects.

This account of the initial causes of variations rests on several assumptions about individual agents. First, although they are not equipped with a view of the operation of the social order as a whole, they must be to some degree both self-aware and situation-aware. Moreover, they do not act on established rules entirely uncritically. They appreciate that the rules make a demand on them, which they ignore only at some cost, but they are sometimes willing to risk paying those costs in order to better realize their goals. This judgment of there being a better chance of realizing their goals may be limited to a single rule that seems to stand in their way, but, since rules come in complex packages, it may also involve a more complex assessment that includes awareness of the way that rules interact in particular circumstances to limit or expand opportunities for successful realization of goals. That is, these characters are norm-appreciating, norm-following, self-aware and situation-aware local optimizers (or satisficers), who may also be aware to some extent of how the system of rules to which they are subject work together to structure the situations and options they face. Hayek's language of "trial and error" is vague, but he must have something along these lines in mind.[29]

How then are these modifications established as rules for the group? We learn from each other by example and imitation, Hayek argues, although neither the teacher nor the pupil may be able to articulate the rule they observe [Hayek 1960, 28-9; Hayek 1973, 19; 1977, 166]. Although he insists that learning by experience is "not primarily by reasoning, but rather by observance, spreading, transmission, and development of practices" which prove successful [Hayek 1973, 18], nevertheless, it is not in mere (unquestioning) *observance*, but in *observation* of the rule-directed behavior and its local success, that imitation is rooted [Hayek 1960, 28]. Imitation starts small, but through the accumulation of large numbers of such small deviations yielding individual rules ("practices"), which then catch on with others, the rules eventually spread through the group to a point sufficient for them to be established as a group practice and have some influence on the social order as a whole. Moreover, since the rules in question have been taken up and used by lots of people they prove to be serviceable in a wide variety of circumstances [Hayek 1976, 21; Heath 1992, 42]. Not all individual

[29] I am here drawing out implications of Hayek's vague language of trial-and-error-generated changes in the rules. Hayek frequently claims that this process is unintentional and blind relative to larger purposes and aims. If he means by this that changes in individual rule-following behavior *happen* but the changes are *not made* by the individual, then we must conclude that Hayek has no account of the mechanism of change and his model of social evolution is fatally incomplete. On Hayek's model, mere changes in behavior are not directly selected by evolutionary forces, because those forces operate directly on social orders. The changes in behavior become relevant to evolution only when they congeal into rules which are replicated in the decisions and actions of a number of members of the group sufficient to effect change of the social order. Thus, to save Hayek's account, we must take his talk of the unintentional and blind character of the process to refer to an individual's lack of awareness of systemic effects and purposes at the level of the social order, leaving him space to develop an account of micro-level intentional activities of individuals along the lines suggested above.

rules are imitated, not all imitated rules spread, and not all rules that spread get established in the group as a whole, but some do and among those that do some will introduce changes in the social order that better equip the group to meet its challenges. Of course, some rule-changes may make the group less effective, and in that case rules that catch fire may die with the group that practices them.

Problems of Identification and Normativity: the Possibility of Common Social Rules

Hayek's explanatory theory, integrating two complementary explanatory schemata, is impressive in its scope and ambition, if disappointing in its lack of rigorously articulated detail. One may wish to challenge the theory at several points, but I propose to inspect just one aspect of process of emergence of rules on which both schemata depend: the mechanism of change in the evolutionary story which is also the pivot of the equilibrating mechanism of the spontaneous order story. Hayek's discussion at this point is critical for the success of his explanatory account as a whole; it is also the point at which his theory joins issue with recent attempts to explain the nature and dynamics of social rules, customs, and conventions. Let us then take a closer look at Hayek's account of the process of rule-change.

The Tasks

To begin, it is useful to note an important difference between biological and social evolution. Biological evolution is *endosomatic,* as Popper would put it; that is, it proceeds by selecting species physical traits that are expressions of genes. As we have seen, in Hayek's account of social evolution, rules play the role of genes. Social evolution is quasi-endosomatic (or, if we tolerate neologisms, *endopsychic*). I say "quasi," because the rules are rooted in subjective dispositions of thought and action (which, of course, supervene on a somatic base). Hayek's story of social evolution is a story of rule-formation, rule-transformation, rule-transmission, and group rule-adoption. This story introduces a level of complexity and a set of problems not encountered in biological evolution, for what must be explained is the emergence and establishment of rules in the behavior of a group. There are at least four interrelated but distinguishable tasks: to explain (i) how it is that *rules* emerge, which (ii) are *social* rules and (iii) the *same* rules across individuals, which (iv) then *spread* through the group as a whole. Let us look at each of these tasks.

First, Hayek must explain how it is that *rules* emerge and change through the activities of an individual's "mistakes" or "experimentation." Rules that allow for change are flexible because they are "voluntary." But this flexibility must be of a certain kind. The pattern-consistent behavior now in view is *called for*, not merely produced by, the rules, and off-pattern behavior must be understood to be a *violation* of the rule, not merely a deviation from the pattern. That is to say, the rules now in view have an essential normative dimension. Thus, for an individual to grasp the rule, it is not enough that she behave in a rule-consistent way; she must also grasp it *as a rule*. This does not require, of course, that she be able to articulate this recognition, let alone be able to explain its rationale, but it does require that in her practice of the rule she understands both that it is possible to act off-pattern (that her compliance is to that extent voluntary) and that off-pattern behavior is not merely *deviation* but *deviance*—that it is not merely different from what the rule would lead one to expect, but that it *fails* to conform to the rule. For this, the "flexibility" required of the rules is two-fold: deviation must be possible and the individual must have some degree of distance from her disposition, such that the option of deviance is open to her. This distance is even more important if she is to relate the consequences of her deviance from the rule to the rule itself in her "trial and error" experimentation. Indeed, on Hayek's model, the individual must be aware of the rule and its suitability for the situation she faces—understanding that situation in terms sufficiently general for her relating consequences back to the pattern which also applies to other situations—and the place of the rule in the complex of rules that give each rule its meaning.

In addition, individuals must be able to recognize the difference between mere *non-conformity* with a rule and *conformity* to a *different* rule. This generates two problems or tasks. First, if new rules are to be introduced by the behavior of the individual, if only on a trial basis, the individual must have a motive to look to an alternative rule, rather then merely to exploit opportunities for improvement of his condition within the regime of existing rules. Hume's "sensible knave" experiments with various forms of conformity and non-conformity with existing rules, but has no interest in introducing new

rules into the regime. He aims only to take advantage of the convenient cooperation of others. Hayek's account needs some answer to the challenge posed by the sensible knave. Second, the rule-tester must have some means of "enacting" or putting in place an alternative rule by means of his deviant behavior. To put the problem in terms familiar to students of customary international law, the question is how is it that *ex iniuria oritur lex*?—how can deviance create new rules? In Hayek's framework, the answer to this must start with the actions or attitudes of the individual rule-innovator, they must choose to "enact" a rule for themselves, and it continues with an account of how this innovation is taken up by others. Unfortunately, Hayek is silent on the initiation of the process. But assuming some account of this process we must consider the problem of up-take, which has two dimensions: how to establish rules that are rules *for* the group (social rules) and rules *in and of* the group (taken up by the group).

Thus, Hayek faces the task of giving account of the rules as *social rules*; that is, he must explain how rules emerge, not just *personal* rules for the individual who is "testing" the rules, but rules *for* the group. This is necessary because the rules he seeks to explain are rules coordinating complex interactions among agents whose actions are interdependent—the outcomes of the actions of each are the vector sum of the actions of all in a context of limited space, time, and resources—and who must be aware of this fact and of the fact that others are aware of this. That is to say, they must be, at a minimum, strategically rational. This awareness, we must assume, is available to individual rule-innovators because they are aware of their local situations and these features are impressive, salient features of those situations. So, in circumstances characterized by a high degree of interdependence and the persistence of coordination and cooperation problems, the individual must not view the world as exogenously determined parameters for his own decision, but rather he must look at the whole system of interactions and consider a rule, or a number of subtly interconnected rules, which all those involved in the concrete problem of interaction can jointly follow. In Hume's vivid image, the task for rule-innovating individuals is not that of replacing one brick in a wall with another, but rather replacing a stone in an arch or vault, each stone of which depends on all the others for the stability and integrity of the vault [Hume 1998b].[30]

The task for Hayek's evolutionary account at this point is to explain how individual rule-testers identify rules of the kind that can perform this complex social function. Moreover, no adequate "test" of such a rule can be performed unilaterally. The rule must be taken up to some degree by others in the group facing the common interaction problem. As Lon Fuller pointed out years ago [Fuller 1969, 4], getting customs or conventions started in conditions of complex social interaction is not like blazing a path through the undergrowth, each successive party treading the path making it more distinct and less formidable. Establishing the rule itself requires coordination.

This brings us to the third task such an account of the emergence of social rules faces. The rules in question must not only be rules *for* a group, they must be rules *in,* practiced by, the group (or some sub-group large enough to provide a good test of the rule). That is to say, the rules must be passed on, transmitted to others. Hayek's proposal here is that rules are transmitted by "imitation." The suggestion is that an individual enacts a rule in his own behavior (however that is accomplished) and this rule is observed, and its example is followed, by others. For this to happen, the observer must recognize the rule-following behavior of the innovator. Only some behavior responsive to observing the rule-innovative behavior of another agent will result in transmitting the new rule. This problem has two dimensions. First, what must be observed and imitated is not merely deviation from the established rules—the sensible knave's exploitation of the cooperation of others—but rather behavior conforming to an alternative rule. "Do as I do" in cases of deviation from established rules is crucially ambiguous between these two modes of imitation. This is not a problem of motivation, but rather a

[30] In game theoretic terms, what the individual seeks is some device that yields a "correlated equilibrium" [see Vanderschraaf 1995; Postema 1998]. The rule, "go on green, stop on red" at intersections that have red-green traffic signals is such a rule, as is the rule "yield to traffic approaching the intersection from the right, otherwise proceed." Note that these rules call for different behavior from, i.e., they assign different "roles" to, the parties approaching an intersection depending on their relationship to some external feature of their common situation (the traffic signal, e.g., or the spatial relations of the parties).

problem of interpretation of the "example" set by the observed behavior. Second, if the example is regarded as an example of alternative-rule-following, the pressing problem for the observer is determining what that rule is. What is needed for rule-transmission, on analogy with reproduction of new individuals with mutated genes in the biological case, is that *the same* rule is passed on. This requires that the rule in question be identified. This *problem of identification* bedevils almost all current game theoretic accounts of the evolution of social rules, although the problem is systematically masked by theorists,[31] and it poses a key task for Hayek's theory as well. Imitation may be involved, but we need some reason to think that imitation is a reasonably faithful reproducer of the rules from the rule-introducing member of a group to members.

Finally, since on Hayek's theory rule-innovation begins in small local contexts but, given the nature of spontaneous order, new rules can only influence the social order as a whole if they are practiced widely in the group, he needs an account of how such rule-innovations *spread* through a population. What must be explained is both how the new rules spread and how the changed rules that spread are relatively faithfully reproduced across the larger group population. This magnifies the problems mentioned in the previous paragraph. The mechanism of change must be capable of producing and reproducing social rules in the group with a substantial degree of fidelity, otherwise rule-innovation will only be a cause of noise, disequilibrium, and eventual deterioration of the established rules. Hayek must explain why we can hope that out of the process of individual rule-testing, new regimes of rules *of the group* can emerge.

Does Hayek's Theory Permit Successful Performance of these Tasks?

Hayek must solve the above four problems for both his theory of spontaneous order and his evolutionary theory to succeed. They must be solved in order to make plausible his claim that a spontaneous order is self-maintaining and his claim that sometimes disequilibrating forces establish new "orders of action" rather than merely set off a spiral of disorder, which can then be tested on the field of group competition where evolutionary forces work. Of course, for his account to succeed, it is not necessary that all rule-innovative activities of individuals result in establishing new group-wide regimes of rules and corresponding social orders. It will surely be the case that some individual "experiments" amount only to knavish deviance without establishing new rules, and some individual rule-innovations will not be taken up by others or not taken up by a sufficiently large sub-population that the social order is materially affected, and some such group-wide rule-innovations may produce overall disorder rather than new forms of social order. Nevertheless, Hayek must be able to show how each of the above problems can be solved and give us some reason to think that the solutions achieved at each level are frequent enough to provide the mechanism of change that the spontaneous order and evolution explanations rely on. If "solutions" are only random and rare, the proposed explanatory schemata fail. They will not be able to account for the actual emergence, existence, and operation of social rules as it promises.

Hayek's understanding of rules and the agents directed by them to some degree promotes solutions to these problems, but it also puts substantial obstacles in the way of solving these problems. First, his understanding of rules as *dispositions* of thought and action is too limited to permit room for the normative dimension of the rules he seeks to explain. Glass has the disposition to shatter when struck sharply, but should a pane of glass fail to shatter upon being struck sharply, no *violation* of a rule has occurred, only a deviation from the expected pattern of glass-shattering behavior. The deviation calls for a revision of our understanding of the disposition, not for change in the behavior of the glass. This familiar point needs no further elaboration here. The consequence, however, is that the notion of dispositions alone cannot provide the conceptual resources needed to explain normative rules. Equipped only with the notion of dispositions we cannot distinguish deviance from deviations, let alone distinguish conformity to a new rule from mere non-conformity. That is, Hayek's very broad and undiscriminating understanding of rules seems to preclude recognition of their dimension of normativity.

[31] For a discussion of this problem see Gopal and Janssen 1996 and Sugden 1998a.

This creates an obstacle to successful performance of the tasks outlined above, however, let us assume Hayek's individuals have the capacity to grasp rules as norms for their behavior and ask whether Hayek has an answer to the knave's challenge. Here the prospects are a bit brighter. The knave's challenge can be seen as two-fold. (i) Given the distinction between mere non-conformity and conformity to an alternative rule, the individual rule-tester must have some motivation to test alternative rules, rather than merely seek opportunities for advantageous non-conformity; and (ii) this alternative behavior must be recognized by others as alternative-rule-following rather than mere advantage-seeking non-conformity. Hayek's response to the first of these challenges rests on his rejection of the conception of human rational agents from which the knave's challenge seems to emerge. Human beings are, first of all, not rational expected utility maximizers, but rather *rule-following animals,* he insists [Hayek 1973, 11]. Of course, they seek to satisfy their desires and realize their aims, but they always do so within a framework of rules. This ordered structure is important to them because the rules provide a source of *intelligibility* in their social lives, which, presumably, Hayek takes to be of more fundamental to them than marginal gains from exploiting free-rider opportunities.[32] Thus, rules, for Hayek, must not be conceived as obstacles to achieving maximal utility (more or less useful under some circumstances), but preconditions for effective or "successful" pursuit of ends, and, it must be said, preconditions of having or forming meaningful ends in the first place. In view of this importance to the individual of being able to see his own behavior as rule-governed as well as to see his environment as intelligibly ordered, the individual will not be willing to adopt strategies that significantly risk disorder either at a personal or social level. Thus, while within limits the individual will have an incentive to explore adjustments of the rules to enable him better to realize his aims, he will do so with a keen sense of the need for an ordered structure for this pursuit, both for himself and for others.

Hayek also seems to have resources for answering the second part of the knave's challenge. Again, because of the importance of intelligibility and rule-governed order, individuals, we can assume, will be primed to recognize such behavior in others; moreover, Hayek argues, our foundational mimicking skills make it possible for us to recognize such behavior in others. So, we can conclude that if he has resources for explaining norm-grasping capacities of human beings, Hayek can solve problems posed by the knave.

However, we cannot be as sanguine about his ability to solve the problems posed by the need for common social rules. The device of imitation alone cannot be sufficient to establish *social* rules needed to coordinate behavior in a systematic way, because it is not possible unilaterally to manifest such rules which can then be imitated by others. The problem at this point is that Hayek's story is at least incomplete. We need a richer account of the capabilities and resources on which individuals can draw to grasp and seek to solve problems of complex social interaction. Hayek, however, is reluctant to do so, because that might seem to require of the agent more awareness of the larger systemic situation, and especially of the complex substratum on which rules depend, than he is willing to allow.

His doctrine of the inaccessibility of the substratum of rules of thought and action puts a far more serious obstacle in the way of successful explanation of the emergence of common social rules. This doctrine creates difficulties at two points. First, it makes it very difficult to see how individual rule-innovators can achieve the distance on the rules that direct their behavior that is needed to assess the rules. Second, because this substratum is not only inaccessible to each individual, but also *private*, Hayek is unable to provide a solution to the problem of identification. Hayek allows that there may be some degree of overlap of basic rules of thought and conduct among individuals in a group, since they will have encountered largely the same natural environment and since they will have picked up many of the same behavioral routines through primitive mimicking of the behavior of others in their group. These similarities may fund, to some degree, reliable expectations regarding the regular behavior of others. But Hayek insists that these commonalities are very limited, restricted to routinely occurring circumstances. But novel circumstances and problems of interaction arise constantly and, he believes, we lack the capacities and resources to resolve them spontaneously. This is due, in large part, to the

[32] Hayek 1967, 90-91. Hume also seems to have advanced this sort of argument in his reply to the sensible knave [Hume 1998b]; or so, at least, I have argued Postema, 1988.

inaccessibility to ourselves, and hence to others, of the vast substratum of experience and knowledge on which our rules are based. Hayek was not forced to this conclusion; indeed, I think he had the resources to explain how a relatively rich "commons of the mind"[33] might develop from the exercise of innate capacities for mimicking and sympathy (as Hume called it) in the thick social environments in which human beings develop, but Hayek refuses to take this route, emphasizing, rather, the inaccessibility and privacy of experience.

It seems, then, Hayek's explanatory project runs aground as a result of two key problems: he cannot account for the normativity of social rules and he cannot solve the problem of identification. Because he cannot explain how those observing the behavior of rule-innovators hit upon *the same* rule, he cannot explain how rule changes introduced by individuals are reproduced and spread in the group. The forces of change which drive both spontaneous order and the evolutionary process either grind to a halt or offer no hope that the result of individual rule-innovative activity will not lead predominantly to undermining of social order.

But this conclusion may be too hasty. In fact, Hayek offers an explanation of normativity of social rules which may also enable him to explain how the problem of identification is solved in social groups. Normativity, he maintains, is a dimension of only some rules of thought and action [Hayek 1973, 43, 74-75]. Normative rules emerge when individual intellects begin to differ in their perceptions or conduct and there is a felt need to reconcile the differences and to teach and enforce the rules. On this view, appreciation of normativity emerges when individuals observe the possibility of deviations and come to appreciate the need to treat them as violations to be corrected. This much, while rough, is not implausible, but then Hayek's thought takes a surprising turn. Because with the emergence of normative rules comes a felt need for reconciliation of differences regarding the rules, Hayek maintains that this task must be assigned to some agent who can resolve the difference (since members of the group cannot do so on their own). They are, he maintains, assigned to "chiefs," judges, and other authorities who *articulate* the rules and *impose* them on the group [Hayek 1973, 43, 45, 77-78]. They express the rules in a form that can then be communicated and explicitly taught and they call upon members of the community to comply with them, backing up their demands with appropriate sanctions. Authorities are not empowered to make any rules they please, he argues, but only to fill gaps in the body of rules already established (in the most primitive instances, presumably, by natural processes). Expectations are shaped and naturally coordinated by implicit common rules and it is the job of authorities to maintain as best they can this structure of coordinated expectations [Hayek 1973, 99-100]. Thus, they are called upon to fit their newly articulated rules into the framework of rules already in place, with a view to the system of rules and the resulting order of actions as a whole it makes possible. Their aim, Hayek insists, is to maintain the proper functioning of this order of actions, and the measure of its proper functioning is that satisfaction of legitimate expectations is optimized [Hayek 1973, 86-87, 99-103, 116].

However plausible this story may be as an account of (a certain form of) common law reasoning on which he models this part of his account, it surely cannot help him solve the problems threatening to undermine his spontaneous-order-cum-evolution account of social rules. For at the point of introducing authorities empowered to manage the system of rules and maintain the order of action, we have left behind all efforts at explaining the emergence of social rules out of spontaneous, impersonal, and unintentional processes. Authorities, as Hayek describes them, impose explicitly articulated rules where the naturally generated rules run into a swamp of diversity and they do so with the proper functioning of the whole system of rules and the social order they produce fully and explicitly in mind. The order, of course, is thought to have no specific goal other than that of coordinating the expectations of members of the group, but that makes their perspective no less systemic and comprehensive, and their efforts to maintain it no less intentional and "planned." Thus, it is fair to say that Hayek's only developed reply to the problems of normativity and identification is one that does not rescue his favored scheme of explanation, but abandons it. Or at the very least, he must concede that *no* social order is entirely spontaneous and the evolution of social rules is from the beginning assisted by intentional, system-aware and group-oriented agents of innovation.

[33]This is Annette Baier's [1997] felicitous phrase.

How might we try to solve the problems of normativity and identification still looming for Hayek? At this point I propose we turn, finally, to Hayek's intellectual friend, Karl Popper, who may have some resources to offer towards a solution to these problems, although for Hayek they may come at a rather high philosophical price.

Objectivity, Discursive Capacities, and the Evolution of Social Rules

The Objective World and Discursive Reason

Despite his fundamental subjectivism, Hayek seeks to assure us of the objectivity of social rules and the judgments we make on the basis of them. His basic idea is that social rules are objective in the sense that they are reliably connected to the world outside the subject [Feser 2006, 304-6]. They are *connected to* (but not, as far as subjects can tell, *reflective* or *true of* that world) by virtue of being *adapted to* that external world. He is also, from time to time, inclined to infer further that we have good reason to rely on them even if the rationale for them is unavailable to us; we have reason to accept them "uncritically," as Hayek often puts it. We have seen already that he is not entitled to these conclusions, without substantial additional normative premises.

But we might think that objectivity is not a core concern for Hayek. His spontaneous-order-cum-evolution schema is meant to be explanatory. The explanation must be illuminating, but it need not for that purpose assure us of the objectivity of social rules. For Karl Popper, in contrast, the idea of objectivity is central to his explanatory project. Indeed, it is only with the emergence of what he calls the objective world (or "world three") that social norms, and human reason itself, become possible. To understand this strange idea we need to survey briefly the nature and evolution of the province of objectivity.

In addition to the inner world of subjects—the domain of mental states and dispositions ("world two")—and the world of physical objects ("world one"), there is, according to Popper, a domain of intelligibles, of logical and cultural objects that he calls "world three." Its denizens include the contents of thoughts (the objects of thinkings), numbers, theories, conjectures and hypotheses, arguments, problematics and unsolved problems, as well as cultural things with physical (world one) dimensions like tools, buildings, sculptures, plays, symphonies, and, most importantly, language [Popper 1972, 106-7; 1994, 5-6]. This world is *objective* in two respects. First, it is objective in the sense that, although its denizens are products of human activity, they do not depend for their continued existence on their makers [Popper 1972, 112]. These objects are not mind-dependent. Popper rejects any form of psychologism or subjectivism that seeks to reduce world three items to items in the consciousness of subjects. World three is a separate domain. It is brought into being and continually added to by human subjects, but beyond that it is not dependent on them. This domain is also objective in the sense that it is *autonomous*. New items in this domain emerge *on their own* from other items in the domain. Arguments may have consequences yet unrecognized by those who entertain them; new problems, new possibilities for argument, and yet unexplored implications and presuppositions of thoughts are regularly generated in this domain [Popper 1972, 117, 159-61; 1994, 19-20, 24-46]. The natural sequence of numbers, Popper maintains in a favorite example, was a human creation, but once taking its place in the objective domain it generated its own problems, which then were available to be discovered by other thinking subjects [Popper 1972, 117]. Thus, this domain is full of unintended and as yet unappreciated consequences of human invention [Popper 1972, 159-60; 1994, 26]. The autonomy of items in this domain and their impact on subjects, for Popper, strongly argues for their independence from subjective mind.

As mentioned, the objects of this world are, directly or indirectly, natural and unintended products of the human activity [Popper 1972, 112]. Other animals also participate in populating this world—beavers construct dams, spiders spin webs, many species use a form of language to express interior states or communicate information [Popper 1972, 115; 1994, 82-3]—but the world three village became a mighty nation after the evolution of human language. Not all the items in this domain are linguistic or admit of linguistic articulation (music, for example), of course, but Popper maintains that language lies at the foundation of the domain and it is the primary medium in which it grows [Popper 1994, 34, 38, 81]. The special and especially fecund genius of human language lies in its two

"higher" functions. In addition to expressive and communicative or signaling dimensions shared with animal languages, there evolved with the human species language capacities that enabled individuals to describe the world around them. The *descriptive* resources of language funded the possibility of offering descriptions that did not match the world experienced. This spurred two key developments: (i) the development of imaginative capacities of the human mind—subjects could frame thoughts of counterfactual situations and invent stories—and (ii) the development of criteria for assessing descriptions as true or false and other logical operations on descriptions. These function as regulative ideas governing the activity of describing [Popper 1972, 119-20; 1994, 81, 86-7]. And this, in turn, made possible the development of the *critical* or *argumentative* use of language. For with tools for assessing descriptions came critical assessment of proposed descriptions and explanations, and with them came resources for evaluating these criticisms, regulative ideas of validity and soundness of arguments, relevance of evidence, and the like [Popper 1972, 119-20; 1994, 86-92]. Rational criticism, working on the offerings of imagination, then became the main instrument of growth of knowledge (i.e., of world three) and evolution of the human species.

The emergence of these dimensions of language, natural products of human activity, gave rise to new human capacities for imaginative and critical use of language and with them a practice of conceiving and critically assessing proposals for making their environment and experience intelligible. The three worlds are distinct, but they interact in important ways, with the second world mediating relations between the first and third worlds [Popper 1972, 112, 117, 147; 1994 7, 20-21]. Human minds think up solutions to problems created by interaction with the physical world and create physical objects—e.g., tools, buildings, ways of producing food—to solve those problems; in turn developments in world three act on the subjective conditions (mental capacities and dispositions) of its creators. "The human mind evolved together with world three," Popper insists [Popper 1994, 10]. Rationality and the self, our rational capacities and practices, and each individual's sense of self, are developed by engaging in rational criticism. Through participation in rational critical activities made possible by the objective domain self-transcendence is possible [Popper 1972, 147-8; 1994, 130-40]. We come equipped, genetically or through social learning, with dispositions, routines of perception and action, and expectations about the world around us, but by bringing those dispositions, routines, and expectations into the light of the objective world, we can subject them to critical assessment. Thus, we are never wholly prisoners of our local environments or instinctive routines, dispositions or expectations [Popper 1994, 139]. As he is fond of saying, we are able to throw a rope into the air and scramble up it, provided it gets a hold in the world of critical discussion [Popper 1972, 148].

Moreover, because the domain in which rational criticism takes place is not private or subjective, but rather objective and in that sense public, and it is always available to rational minds, rational criticism is always capable of being *intersubjective*. Critical argument is, typically and in its primary form, *discoursive* (to use the now obsolete but extremely useful term); that is, it is a matter of offering reasons and arguments to engaged interlocutors. Self-criticism is possible, of course, but it is tutored in a practice of mutual criticism [Popper 1966 vol. 2, 225-7, 238]. World three is the domain where rational subjects meet, subjects whose rationality and "full consciousness" emerge and are nurtured in these interactions, where arguments are explored, new problems discovered, and propositions and proposals ("conjectures") are posted, and then are subjected to rational discoursive criticism.

Popper and Hayek on the Evolution of Social Rules

With this brief sketch of Popper's views in hand, let us return to the question of the emergence and dynamics of social rules and the problems that Hayek's theory seemed unable to solve. The first thing to note is that both philosophers regard the question of the explanation of social rules from an evolutionary point of view. Also, both maintain that rationality and the human mind evolve with the evolution of the natural and especially social world. Popper's rejection of subjectivism, and championing of an objective and autonomous domain in which rational subjects interact, leads him to a very different view of the nature of the evolution of social rules and institutions. Popper's views stand in stark opposition to some fundamental features of Hayek's conceptual framework. While Popper accepts that much of our knowledge is tacit and dispositional, in much the same sense that

Hayek gives this notion, he argues that the resources available to us in the objective domain give us access to this implicit, inborn or socially inbred ("traditional") knowledge and make critical assessment of it possible [Popper 1994, 134-9]. Hayek's mistake, viewed from Popper's point of view, is two-fold. First, Hayek reasons that, since it is impossible to put within any one person's purview (or even that of all of us together) all of what we know implicitly, it is impossible to bring any of it (save a very limited part) to explicit consciousness and the light of critical assessment. Since all of it cannot be accessed, it cannot be accessed at all, and if not accessed then not assessed. This, Popper argues, is just a mistake. He admits that most of our individual ("subjective") knowledge is implicit, and adds that a very large part of objective knowledge (that which resides in world three) is also not in the command of anyone, and he accepts with Hayek that no one can get command of *all* of either sort of rules or knowledge. Nevertheless, he insists, it is not true that much of it must remain inaccessible to us. He writes,

> While our criticism cannot tackle more than one or two problems or theories at a time . . . there is no problem or theory or prejudice or element in our background knowledge that is immune to being made the object of our critical consideration [Popper 1994, 136].

The view that rational argument must always proceed within a framework of assumptions and thus that there will always be a set of assumptions beyond rational assessment, is, he insists, just a myth ("the myth of the framework"). It is not possible to access all assumptions at once, but no assumption is invulnerable to rational criticism. Second, and following on the first problem, Hayek fails to see that this process of rational criticism is not a lonely or private affair. It is, rather, a matter of bringing bits and pieces of our background assumptions, which define in part the horizons of our expectations, to the objective domain for public inspection and critical assessment. In this way, the private and ineffable is given public form and made available for intersubjective assessment.

Hayek's view, of course, is that it just is not possible to articulate rules that are embedded in practice and dispositions. But Popper does not have to deny this to insist against Hayek that critical assessment of that which can be articulated, in a tentative and piecemeal fashion to be sure, is possible and plays an important role in individual development and social evolution. The feedback loop from critical assessment of the contents of thoughts and theories to the practices and dispositions of the subjects who think them enables us to have a degree of critical control of even that which is not articulable. Moreover, Popper insists, through such "error elimination" human evolution has largely proceeded, since the emergence of language. The possibility of exosomatic testing of solutions to problems has proved to have enormous evolutionary advantages for the human species, he argues, because it "allows theories to die in our stead" [Popper 1994, 12]. Thus, in Popper's view, Hayek ignores one of the most important engines of rule-testing. Critical assessment by *discoursive* interaction in a public domain is a tool of enormous power, a tool Hayek's theory refuses to deploy (except to put it in the hands of authorities who do the work of rule-fashioning for us). But this amounts to a kind of myopia, for there are resources available to human individuals facing serious problems of coordination for solving these problems, including problems of identifying common rules, *jointly* and *discursively.*

In the place of intersubjectivity based on a conception of common knowledge understood as nested subjective or private knowledge (of the sort: I believe that he believes that I believe that…) iterated ad indefinitum, Popper offers a model of logical space, a public domain or commons, where individuals can meet, engage in deliberation, come to joint solutions to common problems, drawing on common resources. This model puts at the center of the process of rule-formation and rule-transformation the discursive and critical capacities of members of a group, capacities which are systematically left out of typical evolutionary game theoretic models and ignored by Hayek. On Popper's model, the objective domain is a public place in which we can meet in the hope of working out, discursively and critically, the rules we need for cooperation, a place that does not presuppose already shared values, but rather a place of common argument and deliberation, structured, to be sure, by criteria of validity, soundness, and weight of evidence, but a place where these criteria are also

subject to critical assessment. This offers also an attractive notion or model of objectivity as well. We might call it "objectivity as publicity."[34]

[34] As I have argued in Postema 2000.

References

Baier, Annette. 1997. *The Commons of the Mind.* Chicago: Open Court.

Feser, Edward. 2006. "Hayek the Philosopher of Mind". In *The Cambridge Companion to Hayek,* edited by Edward Feser. Cambridge: Cambridge University Press.

Fuller, Lon L. 1969. "Human Interaction and the Law." *American Journal of Jurisprudence* 14: 1-36.

Gaus, Gerald F. 2006. "Hayek on the Evolution of Society and Mind." In *The Cambridge Companion to Hayek*, edited by Edward Feser. Cambridge: Cambridge University Press, 232-58.

Gaus, Gerald F. 2007. "Social Complexity and Evolved Moral Principles." In *Liberalism, Conservatism, and Hayek's Idea of Spontaneous Order*. New York: Palgrave Macmillan, 149-76.

Gopal, Sanjeev and Janssen, Maarten. 1996. "Can We Rationally Learn to Coordinate?" *Theory and Decision* 40: 29-49.

Hayek, Friedrich A. 1952. *The Counter-Revolution of Science*: *Studies on the Abuse of Reason*. Glencoe, IL: Free Press.

Hayek, Friedrich A. 1960. *The Constitution of Liberty*. London: Routledge.

Hayek, Friedrich A. 1967. *Studies in Philosophy, Politics, and Economics*. Chicago: University of Chicago Press.

Hayek, Friedrich A. 1973. *Law, Legislation and Liberty, Volume 1: Rules and Order*. Chicago: University of Chicago Press.

Hayek, Friedrich A. 1976. *Law, Legislation and Liberty, Volume 2: The Mirage of Social Justice*. Chicago: University of Chicago Press.

Heath, Eugene. 1992. "Rules, Function, and the Invisible Hand: An Interpretation of Hayek's Social Theory." *Philosophy of the Social Sciences* 22: 28-45.

Hume, David. 1998a. *A Treatise of Human Nature*. Edited by David Fate Norton and Mary J. Norton. Oxford: Oxford University Press.

Hume, David. 1998b. *An Equiry concerning the Principles of Morals*. Edited by Tom L. Beauchamp, Oxford: Oxford University Press.

Popper, Karl. 1966 *The Open Society and it Enemies*. Princeton: Princeton University Press.

Popper, Karl. 1972. *The Logic of Scientific Discovery.* London: Hutchinson.

Popper, Karl.1979. *Objective Knowledge: An Evolutionary Approach.* Oxford: Oxford University Press.

Popper, Karl. 1989. *Conjectures and Refutations: The Growth of Scientific Knowledge*. London: Basic Books.

Popper, Karl. 1994. *Knowledge and the Mind-Body Problem*. Edited by M.A. Notturno. London: Routledge.

Postema, Gerald J. 1988. "Hume's Answer to the Sensible Knave." *History of Philosophy Quarterly* 5: 23-40.

Postema, Gerald J. 1998. "Conventions at the Foundations of Law." In *The New Palgrave Dictionary of Economics and Law*. Edited by Peter Newman. London: Macmillan, vol. 1,465-72.

Postema, Gerald J. 2000. "Objectivity Fit for Law." In *Objectivity in Morality and Law.* Edited by Brian Leiter.Cambridge: Cambridge University Press, 99-143.

Postema, Gerald J. 2008. "Salience Reasoning," *Topoi,* 27: 41-55.

Ryle, Gilbert. 1949. *The Concept of Mind*. Chicago: University of Chicago Press.

Skyrms, Brian. 1996. *The Evolution of the Social Contract*. Cambridge: Cambridge University Press.

Skyrms, Brian. 2004. *The Stag Hunt and the Evolution of Social Structure* Cambridge: Cambridge University Press.

Sugden, Robert. 1998a. "The Role of Inductive Reasoning in the Evolution of Conventions." *Law and Philosophy* 17: 377-410.

Sugden, Robert. 1998b. "Spontaneous Order." In *The New Palgrave Dictionary of Economics and the Law*. Edited by Peter Newman. London: Macmillan Palgrave, vol. 3, 485-95.

Vanderschraaf, Peter. 1995. "Convention as Correlated Equilibrium". Erkenntis, vol. 42, 1995, pp. 65- 87

Elective Modernism

Harry Collins[♣]

Introduction

It is hard to move on from Post-Modernism because, like other sceptical positions, it is addictive; scepticism can never be wrong. Scepticism has the strength of deduction around it rather than the weakness of induction. The deduction that we can never be certain that the future will be just like the past is always going to be stronger than any induction from past to future. The claim that no one thing is ever exactly like another so that generalisation and classification is perilous, is always going to be stronger than generalisation and classification. The claim that all things are socially constructed, or just plain `constructed,' is always going to be stronger than the claim that just some things are socially constructed. More detailed description is more faithful to reality than less detailed description and hence scepticism is even supported by radical positivism – only momentary sense experiences can be bedrock. And scepticism is doubly attractive because not only is it safe but it is also rebellious: for the sceptic even the most powerful institutions – scientific, religious, financial and political -- are shallow conceits when looked at from the vantage point of what is truly known -- nothing.

Of course, rebellions driven by scepticism find it hard to turn into revolutions because new syntheses are equally vulnerable to the sceptical critique. What scepticism does is level and homogenise. For example, science has come to be thought of as politics by other means, as continuous with religion-based ideas such as intelligent design, no better than the wisdom of ordinary folk, and just one choice among many master narratives or lifestyles.

Because the logic of scepticism is unbeatable it is indeed the case that neither science nor any other cultural choice is forced upon us by the old arguments from truth and efficiency. But a choice has to be made if life is to be lived. Life cannot be lived on the principle of scepticism – even putting food in one's mouth involves trust. And society cannot be organised on the principle of scepticism for a society without principles would be no society. At best, any such society would likely soon come to turn on the principle of physical power. The remarkable thing is that societies can drive themselves with values beyond self-interest – and history, including recent history, shows how tenuous these values are and how rapidly they can be destroyed. That is why the notorious British Prime-Minister, Margaret Thatcher, was so wrong to claim in the 1980s that `there is no such thing as society.' It was an apt slogan for a leader determined to destroy that swathe of delicately balanced values associated with the professions and replace them with self-interest. Her legacy can be seen in the collapse of banks run for the sake of commissions and bonuses in the absence of responsible financial controls to the attempt to replace professional responsibility with quasi-markets in education and health provision. The homogenising tendency of modern science studies, the idea that science is a continuation of politics by other means and the replacement of foundational myths, such as Galileo standing up to the power of Church and State, with a cynical Machiavellianism, is a continuation of Thatcher's work.

Here the main aim is try to begin to reconstruct the professional values of science and to ask what it would be to live in a society in which a central role would be played by the values that underpin the `form-of-life' of science. This echoes the concerns of Max Weber but the main aim is to say how we can maintain such values in the light of what we know today about science that we did not know forty years ago. In the terms of Collins and Evans, 2002, what can we say about science's values in the light of Wave Two's destruction of Wave One.

If Wave Two has shown that arguments that favour scientific values cannot be got from the ideas of truth and efficiency, such values, if they are to inform a society, will simply have to be `chosen'. We can call the basis of a society which chooses such values, `Elective Modernism.' Elective Modernism is, I want to argue, the most attractive successor to Post-Modernism. I want to suggest that Elective Modernism is more appealing as a basis for society than force, religion, or populism. But the choice itself would not be `rational' but more like a moral choice: one would not

[♣] School of Social Sciences, Cardiff University

want to live in a society in which, say, gratuitous torture of the innocent and weak is acceptable even though one could not *prove* it was a bad society. Those who would demand a `proof' of the badness of such a society would have missed the point.

A society based on science is not a new idea but there as many dystopias in the literature as utopias. But the Elective Modernist society would not be based on science, as science was most widely understood up to the end of the 1950s (Wave One of science studies). The sociology of scientific knowledge, and the other sceptical analyses of science that have been carried through over the last decades (Wave Two of science studies), provide an opportunity to rethink what is meant by science and construe the idea of a society based on science in a different way. One immediate difference between what is being suggested here and some of the earlier utopias/dystopias, is the centrality of scientific *values* not scientific *findings*. What is being put forward here is an approach to living in a world based on the search for knowledge not its outcome. The spirit of this exercise is, then, closer to the implicit democratic politics of Merton's norms of science or the explicit politics of Popper's *Open Society and its Enemies* and *Poverty of Historicism*, than to Aldous Huxley's *Brave New World* or John Desmond Bernal's science based socialism.

After Wave Two we know that what we count as scientific values will be still less well-defined than Merton's and Popper's versions even though it will draw from them. But it is the new-found softness of the edges that provides the opportunity to turn to look at the idea of a science-based society in a new way.

Demarcation criteria and forms-of-life

If Elective Modernism is to be a candidate for consideration the `problem of demarcation' has to be solved. What are `scientific values' and how do they differ from religious values, artistic values, and so forth? Every attempt to find demarcation criteria for science has failed because exceptions to the rules have been discovered or it has been shown that the criteria cannot be applied in an unambiguous way. The most nearly successful attempt is, perhaps, Popper's falsifiability criterion. Yet this depends on the establishment of a logical asymmetry between falsification and the flawed logic of induction. We know we cannot prove a scientific law from induction because an indefinite number of observations are needed, but Popper argued that a properly scientific law can be disproved with only one observation. Unfortunately, Lakatos showed, no definite number of observations can disprove a law since each might be a special case (for example, the swan is not really black, it is covered in soot). The asymmetry falls and so Popper's attempt at demarcation cannot be describing the *logic* of scientific discovery.

But actually Popper was right in spite of Lakatos's disproof! And most of the other attempts to construct demarcation criteria which are said to have failed didn't really fail at all. The demarcation problem has simply been misunderstood. Demarcation is not a matter of logic, it is a matter of culture, or `form-of-life.' Why should there be logically inviolable demarcation criteria for science when, as Wittgenstein shows, there is no logically inviolable definition of `game'? As soon as one allows that the process of demarcation is fuzzy – that it is a matter of family resemblance, ill-defined `ways of going on', and socially understood rules rather than logical universals, one sees that science can be demarcated. And we have always known it! In spite of all the sophistication of Lakatos, we have known (at least since Popper) that a claim that can describe ways in which it might be falsified is more satisfactory than an unfalsifiable claim – why not call it more `scientific.' That the difference does not stand up to close logical analysis should be no surprise – no social rule stands up to close logical analysis. What Popper describes is part of what counts as `a proper way of going on' when it comes to science but not to, say, religion; the proper way of going on in the case of religion is to embrace claims with the fervour of absolute truths warranted by revelation. Thus the potential open-endedness of knowledge claims is going to be one of the important features of any Elective Modernism because it is central to the form-of-life of science.

Cultural domains

The notion of `form-of-life' is problematic. There is an argument about what Wittgenstein meant by a `form-of-life' and it may be that what he had in mind was something to do with the physical

constitution of entities such as humans on the one hand and, say, lions on the other. Irrespective of what Wittgenstein really meant, here a `form-of-life' will be taken to mean something that belongs to human societies – the melange of language and `formative action types' that constitute a culture. Cultures are like `fractals.' Games are cultures but football is also a culture and cricket is another culture. Then again, amateur football is one culture while professional football is another. There is no real problem about this – the structure of cultures exhibits itself at different levels embedded within one another – hence `fractal.' Science is also a culture but has embedded within it biology, on the one hand, and physics on the other. What will be called `cultural domains' are at a relatively high level of the fractal structure: they are the level of religion, sport, politics and science. A question for Elective Modernism is the way cultural domains are demarcated one from another by reference to the `formative action types' (or `vocabularies or motive') that make them up. Poppers' falsifiability criterion, as discussed in the last paragraph, is one way in which science is demarcated from religion. The job is to find more of these differences.

Actions and concepts

Perhaps some methodological clarification is needed before the qualities of science as a form-of-life can be set out. Actions are informed by concepts and concepts are made through actions – concepts and actions are `two sides of the same coin.' It has been the claim of the sociology of scientific knowledge that, because concepts and actions are two sides of the same coin, concepts can be understood by studying actions just as much as actions can be understood by studying concepts – that, indeed, was the condition for the existence of a *sociology* of scientific knowledge. But this formulation can be misleading: actions are not infallible indicators of concepts because not all action `tokens' indicate formative action `types.' A form-of-life is constituted through its `formative action types' rather than every action that every actor executes. To give an extreme example, if scientists cheat, as some of them do from time to time, those actions are not constitutive of the scientific form-of-life. If, tomorrow, every scientist began to invent results rather than report measurements, the form-of-life of science as we understand it would cease to exist.

The sociologist, then, is faced with the problem of separating actions that do and do not constitute a form-of-life. The way this problem is usually solved is through some degree of participation in the form-of-life in question (which is also the way `behaviours' are properly assembled into actions). That way the action tokens go together to constitute a formative action type can be separated from action tokens that are not formative. A form-of-life can be properly described in terms of things that actors do, then, only if those actions are understood. This means that even though actions and concepts are two sides of the same coin, the concepts cannot be understood merely by observing the actions from the outside because the actions can be properly understood only if the concepts – the actors' categories – are at least partially understood. In sum, when one says that concepts and actions are but two sides of the same coin, so that one can understand concepts by studying actions just as much as one can understand actions through studying concepts, this does not allow that either can be studied in isolation. Sociologists have to understand concepts to understand actions just as philosophers have to understand actions to understand concepts. Winch (1957) was wrong in saying that sociology is misbegotten epistemology, but Bloor and Collins were equally wrong in saying that Winch could simply be `stood on his head' unless one's notion of sociology includes quite a bit of the conceptual.

When it comes to understanding something at as high a level as cultural domains, most academics already have the necessary degree of immersion in society to understand the concepts needed to understand the actions. And, of course, while philosophers such as Winch were claiming that sociology was misbegotten epistemology, in a similar way they were relying on their default knowledge of the actions typical of different cultural domains. That is why Bloor is correct to say that Wittgenstein is really a sociologist: Wittgenstein's philosophy of ordinary life is as much sociology as philosophy. Thus, both philosophers and sociologists can agree with the proposition in the last sentence of the last paragraph – that if all scientists suddenly started to invent their results then science as we know it would cease to exist. The point is that we already understand science as a cultural domain – it is `science as we know it'!

Now, it is possible that the move, under Wave 2, toward examining the actions of science in ever greater detail, has led the analysts to make mistakes about the very meaning of science. It may be this very process of examination that has led to the notion that all cultural domains are similar. Close examination of the way actors `go on' in science reveals that actors often act within such domains in ways more appropriate to other domains. This happens in an obvious way when, say, a scientist cheats but SSK has shown that scientists *must* draw on actions that are generally taken to constitute non-scientific domains if they are to bring their arguments to a close (for example, if they are to resolve the experimenter's regress). Scientists cannot avoid making political and other `non-scientific' choices if they are to continue to do science. Hence very close observation makes it easy to draw the conclusion that, since much of what goes in science is the same as what goes on in other cultural domains, science is not a distinct cultural domain.

But this conclusion is wrong. The political and non-scientific choices, though without them science could not proceed, are not constitutive of the form-of-life science, which is why we can refer to them as `non-scientific.' This is the point that Collins and Evans (2002) were reaching toward when they made the distinction between intrinsic and extrinsic politics in science. Even though the cultural domain of science is invested with politics, to reach a scientific conclusion by self-conscious reference to a political preference is not to do science. In technical terms, political action is not a formative action type in the case of science.

There is nothing mysterious about this. Think of it this way: scientists have to eat but this doesn't make eating constitutive of science; scientists have to make political choices but political choice is not constitutive of science. Political choice, like eating, is necessary in every cultural domain and that is why it is not constitutive of any cultural domain – the mere existence of political choice cannot demarcate or demonstrate a failure to demarcate. Political choice demarcates only when it is explicit – in that case it demarcates the domain of politics from the domain of science. That is why a scientist, *per scientist*, cannot proclaim that a certain result was chosen because it favoured a certain political outlook and still proclaim that science was being done.

Family resemblance

The criteria that demarcate science will, at best, stake out a group of activities linked by family resemblance rather than the sharp edges of a `set' as in set theory. The idea of family resemblance is a difficult one. It connotes overlap in qualities such that one member of the family can have nothing -- or is it `almost nothing' -- in common with another distant member at the extreme end though there are overlaps between them.

Here is Wittgenstein speaking of the notion in *Philosophical Investigations:*

> #66. Consider for example the proceedings that we call "games". I mean board-games, card-games, ball-games, Olympic games, and so on. What is common to them all? … if you look at them you will not see something that is common to all, but similarities, relationships, and a whole series of them at that. …
> #67. I can think of no better expression to characterize these similarities than "family resemblances"; for the various resemblances between members of a family: build, features, colour of eyes, gait, temperament, etc. etc. overlap and criss-cross in the same way. …
> #68 … I can also use it so that the extension of the concept is not closed by a frontier. And this is how we do use the word "game". For how is the concept of a game bounded? What still counts as a game and what no longer does? Can you give the boundary? No. You can draw one; for none has so far been drawn. (But that never troubled you before when you used the word "game".)

In the last paragraph Wittgenstein points to the familiar point that social rules, including rules of language, are followed without needing to be explicated and that they are open-ended – new instances are continually being invented and affirmed. Given that he also seems to say that members of a family need have nothing in common, it appears that any two things can be thought of as being members of the same family – it is just a matter of working out the overlaps. For example, a rice pudding has a skin, a tennis ball has a skin, so rice puddings belong to the same family as tennis. But such a loose

notion of family resemblance is vacuous – it cannot do the work of demarcating anything – certainly not science.

The answer to this puzzle is to invoke the idea of the `social rule.' Wittgenstein pointed out that rules do not contain the rules of their own application so the recognition of whether a rule is being followed in any particular instance is a matter of social agreement. That is how the meaning of rules evolves – by new instances coming to be agreed to be examples of an existing rule. This means that a rule does not `contain' all its future applications and therefore future applications cannot be predicted – they emerge as life is lived. On the other hand, even though it is not possible to define completely what it is to follow a rule, in nearly all cases it is possible to recognise when a rule has been broken. For example, I cannot define the rule for how close to walk to a person when I pass them on the pavement (sidewalk) and I know that the rule will vary enormously from culture to culture and circumstance to circumstance, yet I know for sure that if I bump up against a stranger of the opposite sex when passing on an otherwise empty pavement I have broken the rule of walking. The way to think about family resemblance when it comes to defining science is like the example of walking. We cannot predefine all examples of what might become counted as science but we can say `this is not science.' Understanding the application of family resemblance is understanding the application of social rules.

To trivialise, I can say that a carrot is not science even though a carrot is linked to science by a set of overlapping qualities: a carrot is orange-coloured; some of wires in the Large Hadron Collider have orange insulation; the Large Hadron Collider is a scientific instrument.[35] The exercise in front of us is to find more relevant examples of what it is to be and not be a science while not being trapped by the logic of exceptions and overlaps.

It has already been suggested that a group that invented experimental results rather than doing experiments and making measurements would be, as it were, `carrot-like.' They would have lots of things in common with the science family – they would, for example, talk of experimental results and write papers in which results were analysed, and so forth – but they still would not be part of the science family. But there are exceptions. Suppose a perfectly respectable scientist accidentally took a drug which caused him or her to invent results without realising it. That person would not cease to be a scientist and if no-one knew what had happened the invented results would become part of science. Or consider a group of sociologists who decided to make up some results and try to publish them as a test of the refereeing process. In that case the making-up of results would be integral to the science. These overlaps and exceptions cause problems only if it is thought that demarcation is a matter of logic rather than understanding social rules. In spite of the overlaps and exceptions it remains clear that making up results is not included in the formative intentions of any science even though one can imagine situations in which making up results would not exclude one from the family of scientist-actors.

The idea of family resemblance is important to the exercise because it warns us not to look for the key to science in some single logical principle. The history of attempts to find philosophically sound demarcation criteria for science has been to invent a key, to see it defeated through its exceptions, and then to invent another potential key, see it defeated and so on. But since we are looking at social `ways of going' on and a family resemblance between a disparate set of activities we should not expect to find a single key but an overlapping family of ill-defined rules. All the attempts to find philosophically sound demarcation criteria that I know of have been sound, even though none of them have proved to be philosophically lasting. All of them point to a social rule and to demarcate science we need to assemble those social rules.

Demarcating science

A conceptual space has now been cleared for the gathering of social demarcation criteria. Most of what will be gathered into the space will be familiar or even prosaic. This is essentially a warehousing operation. We just have to draw up an inventory of what is on the shelves. The only reason we need

[35] I am not quite sure whether this means that every member of the science family has the quality `non-carrotness' so that it is wrong to say that two members of the same family need have nothing in common but perhaps it does not matter.

to catalogue the goods is that they have been scattered by post-modernism (in particular, Wave Two of science studies).

The Importance Of Nominal Problems

Justin Cruickshank[♣]

Introduction

The argument developed in this paper seeks to save Popper from some of his followers and show how Popper's philosophy, or at least a modified version of it, is central to producing an adequate conception of social science research. Popper's impact has, of course, been mainly with quantitative social science, where his methodological prescriptions for testing theories via the hypothetico-deductive (or H-D) method are drawn upon. Others have taken a more flamboyant approach, using Popper's arguments for knowledge being fallible and growing through criticism, to read him as an intellectual and even a political radical (see for instance Fuller 2003 and Sassower 2006). The focus in this paper is on Popper's critical epistemology, rather than his methodological prescriptions for the H-D method, but this does not mean the argument seeks to present Popper as a political radical. Rather, the attempt will be made to show that Popper's conception of knowledge growing through substantive problem-solving is the most useful approach to knowledge growth for the social sciences. In making this argument, the case will be made that Popper's evolutionary epistemology and commitment to methodological nominalism clash with his later turn to realist metaphysics with the latter needing to be abandoned. Whilst there is no attempt to present Popper as a political radical, the argument of this paper does present Popper's thought as progressive, in the sense that there is both an epistemic and moral commitment to increase knowledge through critical dialogue based on substantive problem – solving.

Popper's philosophy is presented here as a solution to the problem of theory in social science. This problem may be expressed as follows. As part of a general rejection of the notions of truth, knowledge, reality and realism, taking place in some quarters of the social sciences, theory too is rejected. This is because theory is taken to be realist, in the sense that it seeks to develop a set of abstractions that decode the real social processes behind the realm of mere appearances (with these being the changing interactions of agents). That is, theory seeks to mirror a domain which may be called the really real domain. It is argued here that whilst the realist notion of theory is untenable, so too are the anti-realist positions of postmodernism and neo-pragmatism because they, rather ironically, end up trading on realist assumptions. As an alternative, it is argued that Popper's work can be drawn upon to develop a more nominalistic and problem-solving approach to theory: here theory can help explain reality without this explanation of reality having to rely on realism. Before developing the case for Popper's critical epistemology, an approach called critical realism will be discussed. Critical realism tries – and fails – to turn assumptions about reality into the ontological definitions that function as the condition of possibility of the natural and social sciences, without making any claim about these assumptions and definitions being conceptual isomorphs of the really real realm behind mere appearances. Critical realism is meant to be a realism which avoids dogmatic speculation about the ultimate nature of reality and, whilst it does this, it is still untenable. Before discussing critical realism and Popper's critical epistemology as responses to this problem of theory, the problem-situation at hand will be described in more detail.

Creativity *Contra* Theory?

The social sciences, especially sociology, have been subject to much disputation concerning the status of the knowledge claims that may be made. Much of this disputation concerned whether such claims should be causal explanations or the understanding of intersubjective meanings. With the rise of postmodernism this dispute broadened out to question the very notions of truth, knowledge, reality, rationality and objectivity. This postmodern challenge can be divided into an optimistic and a pessimistic version.

[♣] Dept. Of Sociology, University of Birmingham, UK

The optimistic version holds that any claim about the world has to recognise not just the instability of language as a medium for describing the word, but also the instability of the social world itself. Such perspectives (see for example Thrift (1995)) celebrate the overcoming of 'essentialism' (meaning the view that one's identity is determined by some fixed biological essence) and the overcoming of the view that identities are determined by homogenous cultures. In place of any emphasis on stability or fixity the focus is on constant change with identities being hybrid mixtures that are subject to reworking. This is celebrated as liberating for the self, because the self is a decentred identity than is not tied to any determinants and, as such, it is free completely to redefine itself. A favoured medium for expressing hybridity is irony and this is clear in postmodern architecture which develops this optimistic approach by playfully mixing modernism with other styles so as to subvert the universalising tendencies of modernism (see Jencks (1996) on this).

The pessimistic version of postmodernism may be said to hold to the hermeneutics of suspicion, meaning that all knowledge claims are taken to be symptoms of an underlying power – knowledge nexus or 'discourse'. The task is then taken to be that of delegitimizing discourses by showing how knowledge claims are not claims about the world that give us truth, but expressions of power. Many who follow Foucault adopt this hermeneutics of suspicion approach and every analysis offered turns on explaining how a discourse operates. One example is provided by Armstrong (1995) who charts the rise to dominance of the medical discourse. He argues that the medical discourse was confined to bodies deemed pathological and sent to medical institutions and that this changed recently with the medical discourse now having power over all bodies. What this means is that now people always police their behaviour to conform to the prevailing medical discourse which defines the normal body as the always at risk body. In other words, medical discourse is not a body of knowledge which liberates us by giving us knowledge about reality but a form of power which makes agency possibly by moulding people to act in particular ways. By describing the medical discourse for what it is, the hope is that it may be de-naturalised and recognised as a nefarious power-knowledge nexus rather than a liberating body of knowledge. As Sayer (2005) argues though, such positions are crypto-normative. What this means is that whilst a normative commitment against the status quo is the driving force for such critique, this cannot be justified, because there is no notion of truth and no notion of any real human essence or real human rights being oppressed. Indeed, any last trace of 'humanism' is rejected outright.

Whilst no one who espouses the sort of positions just sketched out would regard themselves as a realist, one may say that such positions are actually forms of realism. The reason for this is that the optimistic form of postmodernism posits a metaphysics of contingency. To say the world is all in flux and that there can necessarily be no fixity, is not to eschew metaphysics but simply to offer a metaphysics that defines reality as a process of constant change. As regards the pessimistic form of postmodernism and its attendant hermeneutics of suspicion, one may say that this closely parallels the realism that underpinned ideology – critique. For, whilst there may be no recognition of reality or knowledge, it is still the case that the argument trades on a dualism between appearance and reality, with discourses being the 'moving force' beneath the realm of mere appearances. One may try to argue that exposing knowledge claims as symptoms of an underlying discourse is not the same as saying that there are real material structures, such as capitalism, that act as moving forces to control us via ideology. However, whilst it is the case that discourses and notions of ideologies determined by material structures are different from each other, the argument about discourses is nonetheless realist, because it invokes the existence of a stratum of reality (discourse) that has causal repercussions for agents. (One could also point out that many postmodernists are disillusioned Marxists, as Callinicos (1991) argues, who retain the metaphysics of moving forces controlling agents but without the redemptive ending where agents are freed from forces beyond their control.)

So, contrary to any denial of realism, these postmodern positions trade on a realist metaphysic by making reference to the necessity of contingency and the existence of discourses as a moving force. This inconsistency is picked up on by the neo-pragmatist Rorty (1998a and 1998b). He argues that much postmodern literary criticism is a matter of endless 'unmaskings' and that theories about 'discourse' and 'language' offer a new and 'blurrier' object to replace 'history' or 'the working class' as the object fetishised by radical intellectuals. His point is that whatever the arguments about how

postmodernism and post-structuralism differ from Marxism, the arguments advanced are still realist, because they, in effect, seek to gain critical purchase by moving from appearance to reality.

Rorty makes this point as part of a criticism of left wing intellectuals who have turned from substantive social and political problems to 'theory'. Theory in the humanities and social sciences is taken to be the opium of the intellectuals because it tempts them away from the difficult business of engaging with real social and political problems and towards constructing theories to explain the moving force behind mere appearances: theorists are akin the medieval clerics because their special knowledge takes them to a deeper level of reality which mere agents do not understand. With Marxism this was tied to a 'Christian-like' story of redemption in the future, whereas with the optimistic version of postmodernism we are already liberated to revel in our non-essence and, for pessimistic postmodernism, the wait for redemption has been abandoned for the view that liberation is a humanist myth. Theory then is intrinsically connected, by Rorty, to realism, with realist positions seeking some form of dogmatic metaphysical 'one up manship' that trumps other positions by claiming to know the really real realm.

Theorists may be creative in constructing theories but this is not the sort of creativity that is required which, for Rorty, ought to be a creative search for solutions to substantive problems. Other neo-pragmatists also juxtapose theory to creativity (see for example Baert 2005 and Joas 1996), but do so by arguing that theory cannot account for the creativity of agents. For these neo-pragmatists, theory supplies a fixed set of abstract categories which cannot but fail to understand how the social world is constituted by agents in intersubjective networks of meanings who creatively rework their identities and meanings.

For neo-pragmatists then theory is to be rejected because it is intrinsically realist and realism is to be rejected because it is a form of dogmatism that tempts intellectuals to use their creativity in the wrong way and because the emphasis on grand and abstract schemes to explain underlying social processes cannot recognise the creativity of agents. In place of realism, neo-pragmatists subscribe to what may be termed a radical nominalism. What this means is that categories can be freely reworked because they have no determining external referent. To be sure, the notion of a reality beyond ideas is supported, but this reality quickly becomes redundant because it has no role in the free and creative adaption of categories. As Rorty argues (1991: 81), when the die hits the blank something causal happens but there are as many facts produced by this as there are language games to describe it. Given this, there can be no real sense of problems and creativity becomes detached from substantive problem-solving. All we have are freely developed categories which have no limit to their development other than our innate creativity. Any notion of real, objective problems existing independently of our categories disappears altogether. We can create a problem by creatively describing the world in a particular way and we can solve the problem by creatively introducing some new descriptions (for more on this see Calder 2007). Or, to put it another way, we can jump from old to new descriptions, seeking more arresting metaphors (rather than words to represent reality), with this replacing any real notion of a real problem. Hence when Rorty (1998a) discusses feminism, he has to eschew any notion of a 'real essence' being oppressed and argue instead that if feminists find one language game to their dislike, they will need to construct a different one. Any notion of a real problem existing outside descriptions is lost altogether and the problem becomes that of creatively shifting from a disfavoured set of descriptions to a favoured set of descriptions.

As it happens though, neo-pragmatism, like postmodernism, does have a commitment to realism. This occurs with the *theorisation* of the self, which defines the self as being an intrinsically creative entity, with the worst form of harm that could befall the self being that of having its creativity stifled with an identity imposed upon it (Rorty 1992). Creativity then is defined by this theory as the pre-social essence of selfhood which is shared by all people *qua* people (on this see Cruickshank 2003).

So, if theory is to be rejected for being realist and the radical nominalist alternative is to be rejected for being both implicitly realist (in its theorisation of the self) and unable to link creativity to any meaningful notion of real problems, then we have a problem: we can be neither realist theorists nor radical nominalist neo-pragmatists. Two ways out of this impasse will be explored in the rest of this paper. One is a form of realism, known as critical realism, and the other is Popper's critical

epistemology. Critical realists seek to avoid speculation about the ultimate nature of reality (referred to by them as the 'intransitive domain') and, instead, draw out the assumptions about reality that obtain in scientific knowledge, with these assumptions about what reality is being responsible for the success of science. As regards social science, critical realists turn to agents' lay knowledge for ontological assumptions and develop these by linking them to the ontological definitions derived from the natural sciences. This approach to realism seeks to use the definitions of reality as a meta-theory to guide the natural and social sciences. This meta-theory would be vital for intellectual, problem-solving creativity because, for critical realists, successful explanations which solve previous explanatory problems have to be based on a coherent ontology. It is argued here that this approach to realism is still untenable because its attempt to justify its ontological definitions fails and because adherence to this philosophy would preclude the growth of knowledge by precluding the creative development of new theories with new ontological assumptions. Ontological assumptions cannot, it will be argued in this paper, be the condition of possibility of successful science, as critical realists argue and, instead, ontological assumptions change as theories change. In contrast to critical realism's failed attempt to defend a modified form of realism, Popper's critical epistemology, with its commitment to methodological nominalism, does present a tenable way out of the problem – once, that is, Popper's arguments for realist metaphysics have been removed.

Critical Realism: Deriving Ontological Definitions From Exemplary Knowledge

Critical realists seek to develop a philosophy of the natural sciences which is congruent with the history and practice of the natural sciences. This philosophy is then used as the basis for developing a normative approach to the social sciences which will turn the social sciences from immature to mature sciences. In other words, they seek to develop a realist naturalism that fits in with what natural scientists do, rather than trying to impose a philosophical doctrine upon the natural sciences, and which needs to prescribe an approach to social science research which differs from the approaches currently used by social scientists, in order to make the social sciences properly scientific.

Critical realists argue that most positions which seek the unity of method across the sciences are positivist and that positivist methodological prescriptions must fail to account for the way that the sciences gain knowledge. Positivist philosophies cannot but fail to explain how the sciences work because, critical realists argue, they are based on a fallacy identified by Bhaskar and referred to as the 'epistemic fallacy' (Bhaskar 1997: 16). This is the fallacy of 'transposing' ontological questions about what reality is into epistemological questions about how we know reality. In this case, positivism is held to be a form of empiricism, and empiricism holds that knowledge comes from sense – data inputs, so reality has to be defined in terms of fixed empirical regularities that can be directly observed. In other words, positivism has an implicit ontology, which is a 'closed systems ontology' that construes reality as a system of empirical regularities closed to change. For critical realists, both the H-D method and the inductive method are deemed to be positivist. Of course there are differences: with induction one would seek to observe relations of cause and effect to verify a theory, whereas with the H-D method one would be seeking to observe fixed effects produced by causal laws that were unobservable in themselves, so as to corroborate or falsify a theory. Nevertheless, in both cases, testing would be based on direct observation of empirical regularities that were taken to be fixed, that is, both methods presume the existence of a closed systems ontology. This implicit ontology may fit an empiricist epistemology but, for critical realists, it cannot but fail to account for the practice of science because science is based on the assumption that empirical regularities are not fixed but open to change, that is, science is based on an open systems ontology. Cutting the world to fit a theory of knowledge thus misconstrues the world which, in its turn, has to lead to explanatory failure, because, for critical realists, questions about *how* phenomena interact must be based on a correct definition of *what* the phenomena are.

At this juncture two points need to be made about the critical realist argument. The first is that the putative fallacy referred to as the epistemic fallacy is problematic. It is problematic because it is defined so broadly that only an absolutist metaphysical position which sought to define the ultimate nature of reality would avoid it and it is not clear what the actual fallacy is. Any claim about reality which relates to how we know the world rather than the ultimate nature of reality itself, is taken to be

fallacious, but only two examples of this are given by critical realists. One example is positivism and the other example is the relativism taken to be characteristic of postmodernism and post-structuralism. Now if one accepted that positivism and postmodernism were erroneous for construing reality in a way that is different from the way it is defined in science (rather than misconstruing reality itself) and making reality redundant, respectively, then one can still hold that it does not follow that any attempt to define reality through knowledge claims about it is necessarily fallacious. To accept that two approaches to knowledge are fallacious is not to say that any theory of knowledge must be fallacious. This will be pursued later when we see that critical realists themselves manage to fall foul of this putative fallacy.

The second point to note, is that the critique of positivism turns on critical realism developing an alternative rendering of science, rather than simply dealing with the internal logical consistency of positivism. That is, the rejection of positivism for being committed to a closed systems ontology requires critical realism to have already developed an open systems ontology. This could leave critical realists open to the charge that they dogmatically reject positivism because it is simply different from their philosophy of science. However, critical realists would respond by arguing that they do not seek to impose a philosophical doctrine upon the natural sciences but, instead, that they derive their philosophical principles from within the history and actual practice of science. One may describe the stance critical realists take towards science in terms of them treating the natural sciences as a self-justifying epistemic exemplar. The natural sciences may be described as an epistemic exemplar because they have a history of epistemic success, i.e. success in explaining causal processes, and this epistemic success is self-justifying because it is based on the ontological assumptions about the world within the knowledge claims of the natural sciences. In other words, science has produced knowledge without adhering to any form of 'foundational' input from philosophy, and this production of knowledge has not been a happy accident but a result of the correct assumptions about nature within science. These assumptions though are implicit and the task critical realists set themselves is that of explicating these hitherto implicit assumptions and turning them into clear definitions.

This brings us to the distinction in critical realism between the intransitive domain and the transitive domain. The intransitive domain is taken to be reality and the transitive domain is taken to be scientific knowledge about reality. Scientific knowledge is described as the *transitive* domain because scientific knowledge is held to be fallible. The task of philosophy, as far as critical realists are concerned, is that of rendering explicit the hitherto implicit ontological assumptions in the transitive domain and turning these into a clear set of definitions. These ontological assumptions are held to be of vital importance in understanding the epistemic success of science because, for critical realists, ontological assumptions concerning what reality is determine how explanations are constructed. To misconstrue reality means that one will be unable to explain it. Indeed, Bhaskar goes so far as to say that the ontological assumptions derived from the transitive domain are the condition of possibility for science. The reason why philosophy is required to render the hitherto implicit ontological assumptions explicit is that it is assumed that this will assist the progress of the natural sciences by preventing any erroneous explanations being developed. On this view, philosophy is a conceptual 'underlabourer' that can clear away any conceptual confusion over the definition of reality.

The ontological assumptions that critical realist philosophy takes to be implicit in the transitive domain are that the world is a stratified open system. That is, it is open to change at the level of observable events with this change being caused by the interaction of causal mechanisms that are unobservable in themselves; with biological and chemical causal mechanisms being emergent properties that are irreducible down to physics. Theories and methodologies that are concerned with explaining natural reality must therefore seek to explain the operation of unobservable causal mechanisms and give no truck to the notion of relying on fixed empirical regularities.

When it comes to the social sciences, critical realists argue that there are no coherent ontological assumptions. One consequence of this is that the social sciences are, at best, immature sciences. For, without a coherent set of assumptions about what social reality is, the social sciences cannot produce adequate explanations of how phenomena in social reality interact. The task then is to find a non-dogmatic way to arrive at some ontological definitions for the social sciences. One response could be to universalise the assumptions of one existing social science theory but Bhaskar (1998)

rejects this, arguing that it would beg the question. In response to this he turns to lay discourse and treats this as what may be termed an epistemic proto-exemplar. This is because lay knowledge is taken to have true but vague conceptions of social reality in it that philosophy can clarify. These assumptions in lay discourse are that agents have free will but are constrained by social structures. Archer (1995) makes a similar argument, but focuses on lay experience rather than lay discourse. She argues that social theorists have betrayed the insights of lay agents concerning the experience of freedom and constraint by focusing only on structures or agents. These notions of freedom and constraint are taken, by her, to lead to the structure – agency problem and the need to define social reality in terms of social structures interacting with agents.

To define social structures, critical realists construct a contingent naturalism. They argue that social structures may be conceptualised as emergent properties that arise from the actions of individuals but which then become a stratum of reality in their own right that can condition – but not determine – the agency of individuals. These social structures *qua* emergent properties operate in open systems because agents are not passive structural dopes and can change structures. So, both the natural and the social sciences seek to explain the operation of emergent properties in open systems. This naturalism, or unity of method, is a contingent naturalism, because the need to draw upon the natural sciences for a definition of social reality was contingent upon the social sciences having no coherent ontological assumptions and lay agents having true but vague ontological assumptions that were broadly congruent with the ontological assumptions in natural science. Given this, the task of philosophy as regards the social sciences, is to reject previous theories and to argue for all new knowledge claims in the social sciences to be based on the ontological definitions posited by critical realism, if the social sciences are to be mature sciences.

If one adopted this approach then intellectual – scientific – creativity in both the natural and social sciences would be a matter of engaging in substantive empirical research with solutions to explanatory problems being framed in terms of the ontological definitions furnished by critical realism. Creativity here would be underpinned by theory or, to be more precise, a meta-theory that offered some general definitions, of structures, open systems and, for the social sciences, agency. This meta-theory would not legislate on empirical findings about specific research problems and nor would it seek to justify the ontological definitions proffered by saying that they mirror the intransitive domain of reality in itself. Rather, critical realism seeks to assist the creative solution of explanatory problems by supplying some general definitions of reality that are derived from within a self-justifying epistemic exemplar and an epistemic proto-exemplar. The critical realist meta-theory would thus assist the creative solution of explanatory problems in the sciences by ensuring that reality was not misconstrued.

Problems With Critical Realism

Critical realism is an unusual form of realism. Bhaskar (1997: 36) argues that he treats metaphysics as a conceptual science. This is because the emphasis is on explicating the ontological assumptions taken to obtain in the transitive domains: critical realism seeks to explicate the ontological assumptions within the self-justifying epistemic exemplar of the natural sciences and the epistemic proto-exemplar of lay knowledge. What this realist philosophy does not do therefore is try to argue for metaphysical realism or postulate the essence of the really real realm of the intransitive domain. One may say therefore that critical realism takes the linguistic turn because its focus is solely on definitions and the correct use of conceptual language in the natural and social sciences. The role of the philosopher is thus not to speculate about the ultimate essence of reality (i.e. philosophers cannot step outside the transitive domain to define the essential features of the intransitive domain) or put forward methodological prescriptions based on an epistemic theory (as positivism sought to do), but to police the language of the sciences. So, critical realism is talk about talk with the correct talk – the correct use of ontological definitions – being the condition of possibility of the social sciences and a useful way to remind natural scientists of their hitherto implicit assumptions which served as the condition of possibility of the natural sciences.

Taking this approach to the philosophy of the sciences opens up a justificationist problem – situation. We may have moved away from what one may term 'foundational' epistemologies and the need to say how knowledge claims are justified (of course this was not an issue for the alleged positivist Popper). However, one does need to justify the definitions used by critical realism. This brings us to some serious problems though. If we grant that deriving ontological definitions from implicit assumptions within a self-justifying epistemic exemplar (natural science) and from within an epistemic proto-exemplar (lay social agents' knowledge) is valid, then we encounter the problem of the epistemic fallacy, as defined by critical realists. The problem here is that if it is fallacious to transpose ontological questions about what reality is into epistemic questions concerning how we know reality, then critical realism falls foul of this fallacy. This is because the ontological definitions are not taken to define the essential features of the intransitive domain but are derived from within the transitive domain, i.e. the domain of knowledge. So, by critical realist standards, one cannot actually justify the ontological definitions derived from within bodies of knowledge taken to be exemplary or proto-exemplary, because questions of defining reality are translated into questions of how we know reality. To accept critical realism is thus to accept that the ontological definitions postulated by critical realism are fallacious.

The problem with the critical realist construal of the epistemic fallacy is, as noted above, that any discussion of reality which was not absolutist metaphysic that sought to mirror the intransitive domain, would be guilty of this putative fallacy. One could try to defend critical realism by redefining this fallacy to include only foundationalist philosophies which set out to define the object to fit the epistemic subject, such as positivism – empiricism (which does not include Popper's philosophy). This does not save critical realism though because there are two other serious problems which it encounters.

First, as regards social science, the attempt to justify an ontology of structures as emergent properties interacting with agents, by deriving it from lay knowledge, begs the question. To say that one cannot universalise the assumptions of one theoretical perspective because that would beg the question does not mean that one can side-step this problem by universalising the ontological assumptions held to obtain in lay knowledge. To argue that the ontological assumptions of groups A, B and C (with A, B and C being different social scientists) cannot be universalised without begging the question does not mean that one can universalise the ontological assumptions of group D (lay agents) without also begging the question.

One also encounters the problem of begging the question in the way that critical realists make the move from putatively true but vague notions of reality to a formal ontology of structures as emergent properties operating in open systems. The issue here is that if lay knowledge is knowledge of freedom and constraint then this, by itself, tells us nothing more than that agents lack total freedom. One could try to build on this by arguing that individuals were constrained by other individuals; that individuals were constrained by intersubjective meanings rather than emergent properties; or that individuals were constrained by structures *qua* emergent properties, etc. That is to say, the truism that individuals lack total freedom cannot, *by itself*, justify a particular social ontology, without begging the question.

The second problem concerns the philosophy of natural science. If we accept for the sake of argument that there is one set of assumptions in natural science and that critical realism has correctly explicated these, then we encounter a tension between the attempt to answer a transcendental question and the putative commitment to fallibilism. The problem here is that one cannot say that a set of ontological assumptions furnish the condition of possibility of science whilst also saying that science, and hence its ontological assumptions, are fallible. For if one took fallibilism seriously (so it was more than an empty rhetorical gesture), then one would want to address the issue of knowledge claims being revised and replaced. This would presumably mean recognising and being able to account for the change in ontological assumptions about reality that would occur eventually as knowledge in the transitive domain changed. However, if it was argued that one particular set of ontological assumptions constituted the condition of possibility of natural science then one could not allow new ontological assumptions to be drawn upon. The philosopher would, given this, be forced to police a situation of formalised Kuhnian normal science: all the scientists *qua* scientists would have to use one

set of ontological definitions, because alternative ontological definitions and assumptions would be, by definition, non-scientific for the critical realist.

So, this critical realist approach to the issue of theory and intellectual creativity cannot sustain the notion that the critical realist meta-theory is of vital importance to the creative solution of explanatory problems. For, in the natural sciences, it would impose a condition of formalised Kuhnian normal science that would preclude the growth of knowledge and, in the social sciences, the definitions could not be justified. In addition to this, the philosophy fell foul of the epistemic fallacy as constructed by critical realists – a fallacy which critical realists regard as the Achilles' heel of most preceding philosophies.

A Popperian Alternative

Popper's work is rejected by critical realists as a form of positivism because of his advocacy of the H-D method. To rebut this reading of Popper as a positivist one could note that Popper's rejection of justificationism in epistemology led him to reject the notion of a final justification of a refutation (Popper 1994: xxxv). That is, there is no direct or immediate access to reality and thus there is no justification of a refutation based on direct observation of a closed system. Instead, all claims are fallible interpretations of reality where we, to some extent, impose our stamp on the world. Rather than deal with Popper's methodology though, the focus in this section will be on what may be termed Popper's critical epistemology. What this means is that the focus will be on Popper's argument that once epistemology has abandoned the search for justified true belief we ought to conceptualise knowledge as fallible and subject to growth through criticism.

Central to understanding Popper's position is his rejection of 'subjectivist' epistemologies which were concerned with explaining how the epistemic subject can get justified true belief of the objects of knowledge (see especially Popper 1972 and 1974). Popper argued that those philosophies which turned on the subject – object dualism (which are for him Cartesian rationalism together with Bacon's empiricism and the empiricism of the Vienna Circle) put all the focus on the source of knowledge in the mind. The argument here was that the manifest truth could be recognised as such if one paid due heed to the inner source of knowledge, in the form of *a priori* ideas or *a posteriori* ideas. If one failed to do this by, for instance, following social norms, then one was epistemically and morally (Popper notes a religious residue to such positions) responsible for one's ignorance or error. The epistemologies based on this subject – object dualism were meant to be liberating, with the subject having mastery over the object (a point, of course, which postmodernists make much of). However, for Popper, these epistemologies conflate the object into the subject. Focusing on empiricism, Popper argues that defining the world in terms of our experienced ideas of it results in idealism, with the object becoming the idea the subject has of the object.

One way of describing this is to say that Popper anticipated what critical realists call the epistemic fallacy. It is important to note that critical realists regard the identification of this fallacy as an original contribution which is radically at odds with all preceding philosophy. In one sense this is true, because they are original in arguing for the switching of concern from epistemology to metaphysics with metaphysics being defined as a 'conceptual science'. However, in a more important sense, it is erroneous to hold such a view, for many of the debates about idealism and scepticism stemming from the subject – object dualism are, to some degree, holding that it is in error to define the external material object to fit the subject.

What distinguishes Popper is that his critique of the subject – object dualism leads him not to reject epistemology *per se* but epistemology which is concerned with justifying truth claims. One could say that he wants to reject foundationalist epistemology and endorse an anti-foundationalist epistemology which retains the notion of knowledge but which replaces the search for justification with the recognition of fallibilism. Before exploring this notion of fallibilism, we can note that for critical realists Popper's philosophy would still be guilty of the epistemic fallacy. This is because any attempt to develop an epistemology, whether foundational or otherwise, would be committed to the fallacious problem-situation of translating questions about reality into questions about how we know reality. However, the problem with this approach is that unless one opts for an absolutist metaphysics where one defines the really real realm (or intransitive domain), then any philosophical argument

about knowledge and reality will be guilty of this including, as we have seen, critical realism itself. So, rather than define the problem too broadly to reject any form of epistemology, we are better off restricting the problem of the conflation of reality into knowledge to foundationalist epistemology.

Popper's anti-foundational approach to epistemology replaces the subject – object dualism with an evolutionary approach to the growth of knowledge. In taking this approach the problem of reuniting a divorced subject and object to explain how the subject can have justified true beliefs of the object is rendered redundant. In its stead, the problem becomes that of saying how we adapt to the environment that we are always already a part of. This means that the 'passive' or 'spectator' view of knowledge characteristic of foundationalism has to be replaced by an 'active' notion of knowledge acquisition and development: rather than the subject passively receiving ideas of the external object, knowledge is acquired by us interacting with our environment and, specifically, by creating, revising and replacing conceptual tools to do this. So, for Popper, knowledge about our environment is possible and in place of certainty it is fallible because it entails conceptual mediation with reality. Fallibilism is not deemed here a 'second – best' position to be endured, but rather a condition to be embraced, for it is responsible for intellectual and even moral progress. The view here is that as knowledge is fallible it should always be open to criticism and this criticism will result in the growth of knowledge. This critical approach to knowledge growth may be said to result in moral as well as intellectual progress because it is based on and reinforces the liberal values of free speech and tolerance of dissent.

When it comes to applying this to science, Popper (2002a and 2002b) argues for what he terms methodological nominalism and against what he terms methodological essentialism. The latter is characterised by the attempt to base science on definitions of reality, with definitions supplying knowledge by capturing the essential features of reality. By contrast, the former – methodological nominalism – eschews the attempt to explain reality by 'pinning down' its essential features in fixed definitions and, instead, treats concepts as changeable tools that should explain how phenomena interact – not what the really real properties are behind such interactions. This is compatible with the H-D method because whilst the H-D method postulates the existence of unobservable causal laws, such postulations are revised and replaced when predictions based on them are falsified: one may conjecture the existence of a causal law and then replace this with another conjecture when corroborating evidence turns to falsifying evidence. In other words, postulating the existence of unobservable causal laws does not necessarily commit one to the view that claims about such entities may be justified and treated as the essential properties from which we may read off what happens in the realm of empirical observations.

Using Popper's approach we may say that critical realism was a form of methodological essentialism because ontological definitions are taken to be the drivers of intellectual progress. Unlike the methodological essentialism that Popper is concerned with though, critical realism does not posit definitions that are meant to be conceptual isomorphs of the intransitive domain. Nevertheless, critical realism holds that the condition of the possibility of natural science and social science is that they are based on true ontological definitions. This may avoid the subject – object dualism but it still operates within the ambit of the justificationist problem – situation, for critical realists have to justify their definitions and, as we have seen, their justifications for their ontological definitions fail.

Nominalism *Contra* Realism

In his later work, Popper (1972 and 1996) modified his approach to fallibilism by introducing the notion of verisimilitude and he replaced his earlier agnosticism towards metaphysics with an endorsement for realist metaphysics in the form of an argument for metaphysical realism and a position he termed 'modified essentialism'. These alterations are, it will be argued here, highly problematic for his critical epistemology and its commitment to an evolutionary approach to knowledge. Before we explore these problems we need to clarify what these changes were to Popper's philosophy.

As regards verisimilitude, Popper argues that whilst we can never attain absolute truth, this ought to be our goal and, as we pursue this goal, we will get closer to the truth. As regards metaphysical realism, Popper argues that we can neither prove this nor disprove its contrary, which is idealism, because both are metaphysical positions. Nevertheless, he says that there are reasons to

prefer the view that there is a reality that exists independently of our ideas of it, to the view that reality is exhausted by our ideas of it. The basic point he makes is that idealism is an arrogant philosophical conceit that makes reality dependent on us. As Popper puts it '[d]enying realism [and thus affirming idealism] amounts to megalomania (the most widespread occupational disease of the professional philosopher) (1972: 41). He continues by arguing that if realism is true then the reason for the impossibility of proving it is obvious, namely that knowledge consists of fallible or tentative adaptations to reality, meaning that we cannot *justify* any claim about what lies beyond our theories. Nonetheless, Popper argues that we still need to *presume* the truth of realism because, without it, our fallible search for truth becomes pointless (Popper 1972: 41-42). Whereas metaphysical realism simply asserts that there is a reality beyond our representations of it, Popper's arguments for modified essentialism go one step further, to deal with the issue of the essential features of reality. This he describes as follows:

> Although I do not think we can ever describe, by our universal laws, an ultimate essence of the world, I do not doubt that we may seek to probe deeper and deeper into the structure of our world or, as we might say, into properties of the world that are more and more essential, or of greater and greater depth. Every time we proceed to explain some conjectural law or theory by a new conjectural theory of a higher degree of universality, we are discovering more about the world: we are penetrating deeper into its secrets (Popper 1996:137).

So, in contrast to the metaphysical agnosticism of the *Logic of Scientific Discovery* (2002c), later Popper gives us a strong commitment to metaphysical realism and the notion that we are getting closer to the truth about the properties of reality. Two problems with this may be identified.

The first problem to note here is that the argument for metaphysical realism is less than convincing and, as with other arguments which hold that metaphysical realism has to be presumed to make sense of science (see for instance Trigg 1989 and 1993), it begs the question: the view that science can only make sense if one presumes the truth of metaphysical realism only makes sense itself if one is already committed to the view that science must presuppose metaphysical realism. To be sure, saying that science only makes sense if one rejects the idealist view that theories are self-referential, with the world being that which we freely make, and endorses the metaphysical realist view that there is a reality that exists independently of our representations of it, sounds intuitively plausible. However, one does not need to presume the validity of metaphysical realism to argue that theory change is rational because it is a matter of epistemic progress. Indeed, the notion of science having rational theory change is undermined by this metaphysical argument. The reason for this is that it opens up a dualism akin to the subject – object dualism rejected by Popper. In this case, rather than have the lone mind of the epistemic subject divorced from the objects of knowledge, we have our human made theories on the one hand and a postulated metaphysical domain on the other hand which is unknowable in itself. With this bifurcation we can never step outside our theories to see if a theory captures, wholly or partly, the reality that exists independently of our theories. The realm that theories seek to refer to is defined as a metaphysical domain which is beyond knowledge. Of course, metaphysical realism is itself an ontological and not an epistemological doctrine: it does not say whether or not we may know reality (on this see Searle (1995)). Nevertheless, Popper is using this metaphysical doctrine as the condition of possibility of scientific knowledge, by saying that epistemic progress - or, the growth of knowledge – is only possible if one presumes this doctrine. Yet, adopting this position just invites the sceptical rejoinder that what we take to be knowledge not only lacks justification (in the traditional epistemological sense) but that it cannot even be taken as a fallible engagement with reality.

The second problem is that the arguments for verisimilitude and modified essentialism clash with the evolutionary epistemology advocated by Popper. This is because whereas evolution is characterised as a process without a telos or direction, the arguments about getting closer to the truth and penetrating deeper into nature's secrets imply a very clear direction and goal. One may argue that this is a perfectly acceptable position and, if it clashes with other aspects of Popper's work, then those other aspects must be erroneous. There is not the space here to review all the discussions about Popper's evolutionary epistemology but, what we can say, is that these arguments about evolution

towards a certain goal are problematic. For a start one must end up justifying the view that ontological assumptions are becoming progressively more accurate renditions of the really real realm (albeit with this being a never ending process) but no such justification is given. Instead we are told this must be the outcome of the process of problem-solving but this simply begs the question. At this point we may, surprisingly, gain by turning to Kuhn and, specifically, his argument about the parallels between scientific and biological development. Kuhn argues that 'scientific development must be seen as a process driven from behind, not pulled from ahead – as evolution from rather than evolution toward' (2000: 96). In other words, we can improve our conceptual tools by responding creatively to problems but this focus on overcoming problems does not underwrite any notion of problem-solving necessarily producing increasingly accurate ontological assumptions into infinity. And, of course, it does not mean, *contra* Kuhn*,* replacing the notion of problem-solving with the notion of puzzle-solving which is, ironically, more suited to the notion of knowledge evolving towards a particular goal, given that puzzles have solutions. Nevertheless, taking this notion of evolution seriously does mean recognising the existence of a path-dependency, in the sense that theories are not constructed *ex nihilo* but as solutions to past explanatory failures. We are on a path, the direction of which is contingent upon the creative adaption to explanatory failures, and this is not sufficient to presume we are on a never ending path to a God's eye view.

Conclusion: Methodological Nominalism, Problems And Theory

Contrary to Popper's implicit view that the condition of possibility of the natural sciences lies in adherence to a realist metaphysic, we can say that his account of evolutionary epistemology is able to sustain the notion of theory change in the natural sciences being a rational process without the need for such metaphysical support. The reason for this is that Popper's evolutionary approach to knowledge, which holds that knowledge grows through substantive problem-solving, replaces any dualistic conception of the subject, or theories, being separate from reality with the notion of knowledge being always already engaged with reality – knowledge claims, in the form of theories, is an on-going adaption to reality. Given this, one may argue that theory change is a rational process because it is driven by finding solutions to substantive problems. That is, it is a rational process not because it is evolving to the telos of 'deeper' knowledge of a domain defined as separate from knowledge, but because it is a creative response to problems – it is evolving away from past explanatory failure. Central to this is methodological nominalism which construes theories as conceptual tools that need to explain the interaction of phenomena , in contrast to any form of essentialism, which holds that the task of theories is to represent, in a fallible or otherwise way, the essential defining features of reality.

This can be applied to the social sciences as follows. The social sciences, like the natural sciences, do not get knowledge because they adhere to a fixed set of ontological definitions or because general theories are able to map all the essential determinants of social reality. Instead, knowledge grows in the social sciences though substantive problem-solving. This requires intellectual creativity to solve problems and central to this is theory, conceived of in non-realist terms, because it is a tool that we can adapt through our problem-solving engagement with reality. The growth of knowledge in the social sciences may therefore be said to rely on what could be termed nominal problems rather than real*ist* problems: that is, we encounter real problems but these are not failures of theories to represent the really real realm, or failures to conform to a set of ontological assumptions in particular domains of knowledge, but problems of our conceptual tools to deal with the reality that they are already engaged with.

Bibliography

Archer, M.S. 1995. *Realist Social Theory: The Morphogenetic Approach*. Cambridge: Cambridge University Press.
Armstrong, D. 1995. 'The Rise Of Surveillance Medicine', *Sociology Of Health And Illness* 17 (3): 393 - 404.
Baert, P. 2005. *Philosophy Of The Social Sciences: Towards Pragmatism*. Cambridge: Polity.
Bhaskar, R. 1997 (1975). *A Realist Theory Of Science*. London: Verso.
Bhaskar, R. 1998 (1979). *The Possibility Of Naturalism: A Philosophical Critique Of The Contemporary Human Sciences*. 3rd edition. London: Routledge.
Calder, G. 2007. *Rorty's Politics Of Redescription*. Cardiff: University of Wales Press.
Callinicos, A. 1991. *Against Postmodernism: A Marxist Critique*. Oxford: Polity.
Cruickshank, J. 2003. *Realism And Sociology: Anti-Foundationalism, Ontology And Social Research*. London: Routledge.
Fuller, S. 2003. *Kuhn Vs Popper: The Struggle For The Soul Of Science*. Duxford: Icon Books.
Jencks, C. 1996. *What Is Postmodernism?* 4th edition. Chichester: Academy Editions.
Joas, H. 1996. *The Creativity Of Action*. Cambridge: Polity.
Kuhn. T.S. 2000. *The Road since Structure: Philosophical Essays 1970-1993*. London: University Of Chicago Press.
Popper, K. R. 1972. *Objective Knowledge: An Evolutionary Approach*. Oxford: Oxford University Press.
Popper, K.R. 1976 (1963). *Conjectures And Refutations: The Growth Of Scientific Knowledge*. London: Routledge.
Popper, K.R. 1996 (1983). *Realism And The Aim Of Science*. London: Routledge.
Popper, K.R. 2002a (1957). *The Poverty Of Historicism*. London: Routledge.
Popper, K. R. 2002b (1945). *The Open Society And Its Enemies*. London: Routledge.
Popper, R. K. 2002c. (1935). *The Logic Of Scientific Discovery*. London: Routledge.
Rorty, R. 1991. *Objectivity, Relativism And Truth: Philosophical Papers vol. 1*. Cambridge: Cambridge University Press.
Rorty, R. 1992. *Contingency, Irony And Solidarity*. Cambridge: Cambridge University Press.
Rorty, R. 1998a. *Truth And Progress: Philosophical Papers vol. 3*. Cambridge: Cambridge University Press.
Rorty, R. 1998b. *Achieving Our Country: Leftist Thought In Twentieth Century America*. London: Harvard University Press.
Sassower, R. 2006. *Popper's Legacy: Rethinking Politics, Economics And Science*. Stocksfield: Acumen.
Sayer, A. 2005. *The Moral Significance Of Class*. Cambridge: Cambridge University Press.
Searle, J. R. !995. *The Construction Of Social Reality*. London: Allen Lane.
Thrift, N. 2005. *Knowing Capitalism*. London: Sage.
Trigg, R. 1989. *Reality At Risk: A Defence Of Realism In Philosophy And The Sciences*. Hemel Hempstead: Harvester Wheatsheaf.
Trigg, R. 1993. *Rationality And Science: Can Science Explain Everything?* Oxford: Blackwell.

Falsification falsified. A swansong for Lord Karl

Frédéric Vandenberghe[♣]

Karl Popper may well be the most overvalued philosopher of the twentieth century. His neo-positivism did a lot of damage in the natural sciences. As he had to admit that the 'covering law-model' does not really apply to the social sciences, he developed an alternative model of explanation for the human sciences, and introduced the situational logics of rational choice as second best in the human sciences. I will submit critical rationalism to a metatheoretical critique and question its ontological, epistemological, ideological, ethical and anthropological presuppositions from a realist, phenomenological, hermeneutic, communicative, humanist perspective

I am not a polemical thinker. I am not interested in critique for critique's sake. Theoretical critique, however just and justified, is all too often only an easy outlet for free floating anger. If I don't like an author, I just ignore him or her. If I don't like a book, I simply won't write a review. So when I criticize someone's work, it is actually because I value it. Not in spite of the fact that I value it, but because I value it; because I disagree with it, I submit it to critique. I confess that like all of us, I have occasionally committed character assassinations, mostly in footnotes though. But all in all and as a matter of principle, I try to practice the anthropology of admiration, not just in my academic, but also in my personal life.

So why did I write such a nasty first paragraph? Why did I say that Popper may well be the most overvalued philosopher of the twentieth century? Why did I write that he did a lot of damage not only in the natural sciences, but in the human sciences too? Well, I did so, because, as it happens, Popper has succeeded in defending the two very positions that I find intractable: positivism in the natural sciences and rational choice in the human sciences. As a social theorist and a practising humanist, I am willing to contemplate most other positions and I will even incorporate their arguments if I can. Over the years I have studied, learned and benefited from critical theory, structuralism, hermeneutics, pragmatism, phenomenology and ethnomethodology. I do not have much sympathy for psychoanalysis and postmodernism, but if necessary, I'll even bring them in to argue against positivism and rational choice. Mention these and I quickly get grumpy…

But before I launch my critique of Popper, let me pay due honour to the deceased. After all I promised a swansong, didn't I? I admire Popper the teacher, the moralist and the *Aufklärer*. I appreciate the fact that he became a philosopher in spite of himself and that he refused the complacency and smugness of the academic establishment. As an admirer of Socrates, he cultivated the virtues of intellectual modesty and conceptual clarity. He defended his positions with verve, though not always with talent. What I find interesting in Popper are the unresolved tensions in his work. The problems he poses rather than the solutions he proposes. Although he never openly accepted the fact that his theory of falsification had been falsified, he silently revised his own positions. With stubbornness, he defended his irrational faith in reason and accepted the force of the better argument as the only form of force that can be accepted in science and society at large. Towards the end of his life, he anticipated the theories that his critiques would later advance against his earlier work. I'm thinking here above all about his theory of three worlds (which corrects his methodological individualism), his realist theory of propensities (which overcomes his penchant for nominalism) and his defence of indeterminism (which softens his nomological regularism). He may have overestimated his contribution in rebutting the positivism of the Vienna Circle, as Carnap objected[36], he nevertheless should be credited for his trenchant critique of the empiricist dogma of 'immaculate conception'. And although he refused till the end the irrationalism of his conventionalist detractors, not to mention the constructivist sociology of science, both the over- and the underdetermination theses can be found in

♣ Professor of Sociology at the University Research Institute of Rio de Janeiro. I am grateful to the organizers of the conference for their generosity and the secretaries of the Max Weber Program for their patience. I also want to thank Mathias Delori for his trust and Jennifer Greenleaves for her hospitality.

[36] Carnap, R.: *Mein Weg in die Philosophie*, p. 49.

his early work. While the overdetermination thesis states that facts are always overdetermined by theories — in other words that facts are really reified theorems, as Bachelard would say — the underdetermination thesis argues that theories are always underdetermined by facts — in other words that various theories are compatible with a given fact. As Harry Collins has explored all the intricacies and implications of both theses with courage and talent for thirty years, I am only too happy to defer to him and to refer the audience to his work (especially his work on gravitational waves, as well as his experiments with Trevor Pinch in spoonbending).[37]

***

Instead of a hypercritique, I'd like to propose a metacritique of critical rationalism.[38] By means of a metacritique, I want to submit the transcendental and quasi-transcendental presuppositions of Popperianism to philosophical scrutiny. Unlike a straightforward critique, a metacritique does not directly refute a theory. As a form of immanent critique, it takes theories at its words. It thinks through its presuppositions, and uncovering 'performative contradictions', it connects, in good pragmatic fashion, the presuppositions to its consequences.[39] When the consequences are in tension with the presuppositions, as is the case for instance, when an author affirms A, but actually does B, it calls for a dialectical supersession (*Aufhebung*) of the performative tension in a more encompassing framework.
In good philosophical fashion, a typical metacritique analyzes the ontological, epistemological, ideological, ethical and anthropological presuppositions of any given theory. Coming from sociology, I will focus on Popper's philosophy of the social sciences. I will not spare his philosophy of the natural sciences, however. On the contrary. Drawing on critical realism, I will basically argue that the deductive-nomological model does not even hold water in the natural sciences. If that's the case and if the D-N model is indeed displaced in the natural sciences, then I really don't see why one would like to defend it in the social sciences and the humanities.

Ironically, Popper arrived at the same conclusion. Instead of straightforwardly arguing for naturalism and extending the covering law model to the social sciences, he developed an alternative model of explanation and introduced the situational logics of rational choice as second best in the human sciences. Instead of resuscitating Comte's social physics, he has nothing better to offer than social economics. As a standing member of the anti-utilitarian movement in the social sciences, known in France under the acronym of MAUSS, which seeks inspiration in Marcel Mauss's anthropology of the gift, I think, however that this is not good enough. The social sciences should not take their bearings from economics, but rather the reverse. Resisting the colonization of the social sciences by economics, I'd like to reconnect the social sciences to the humanities. The social sciences belong, by nature, to the human sciences. Standing in between the natural sciences (which are descriptive and explanatory) and the humanities (which are interpretative and normative), they partake a bit of both. As such, they develop a third kind of knowledge. It is not theoretical knowledge ('know that', in Ryle's terminology), nor is it merely a craft, a habit, a skill (´know how´). As a reflexive form of common knowledge, yes, as a methodical extension and systematization of common sense (*sensus communis*) that is shared by all who speak the same language, the new sciences represent knowledge of the third type. Perhaps we could call it 'know with' or, in more Peircean vein, knowledge of "withness".

***

Taking the notions of critique and reason seriously, I will now proceed to a metacritique of Popper's philosophy. I will first analyze his contribution to the philosophy of the natural sciences and, then, in a second moment, I will scrutinize his philosophy of the social sciences.

In an attempt to solve the philosophical problem of induction and to demarcate the sciences from the pre-, the para- and the pseudo-sciences, such as astrology, Adlerian psychoanalysis and

---

[37] Cf. Collins, H. M.: *Changing Order. Replication and Induction in Scientific Practice* and Collins, H. and Collins, H. M., & Pinch, T. J.: *Frames of Meaning: The Social Construction of Extraordinary Science.*

[38] On metacritique, cf. Vandenberghe, F.: *A Philosophical history of German Sociology.*

[39] "I was on the boat. The boat sank. Nobody survived" is a typical example of a performative contradiction.

Marxism,[40] Popper has formalized the deductive-nomological model of Mill, Jevons and Whewell and proposed his famous theory of falsification as a normative description of scientific method. While the D-N model states that a scientific theory has to take the form of a universal law from which, together with clearly specified antecedent conditions, a particular event can be deduced, the criterion of demarcation stipulates that a theory is scientific only to the extent that it can be falsified by a test on observable data.

A closer look at the D-N model of explanation reveals, however, that it presupposes the existence of a closed model in which all variables are artificially controlled so as to make predictions possible. As Roy Bhaskar has shown in his path breaking critique of positivism, this only happens in experimental conditions which artificially close the system.[41] Allowing for meticulous control of all the factors and antecedent conditions (which are otherwise smuggled into the *ceteris paribus* clause, which, uncontrolled, creates havoc), experiments make causal explanation and prediction possible. By making abstraction of the causal intervention of the scientist in experiments, Popper has unknowingly identified the laws of nature that scientists observe in experimental circumstances with the laws in nature. The tacit identification of the laws of nature with the experimental conditions in which they can be observed leads to the absurd conclusion that scientists cause and even change the laws of nature!

It is only if the role of the scientist is properly theorized that experiments can be understood as meaningful accomplishments. But paradoxically, that is exactly what Poppers's theory proscribes. By focusing on covering laws and antecedent conditions, the actions of the scientist are reduced to a form of predictable, exodetermined behaviour. As if one could explain human action by taking the collision of billiard balls (preferably white or red) as the model of explanation! At this point, the inhuman consequences of the Humean concept of causality are brought to the fore. In the second round of the *Positivismusstreit* (the first round never took place – Adorno never debated with Popper and Horkheimer never read him), Jürgen Habermas and Karl Otto Apel argued convincingly that the communication among scientists was the blind spot of critical rationalism.[42] By adopting a scientistic perspective on science, Popper had actually curtailed reason and made his own theory immune against experience and common sense. In spite of its insistence on problem solving and the elimination of errors, Popper's neo-positivism is anything but pragmatic. Instead of reconnecting the sciences to common experience, as C.S. Peirce, G.H. Mead and John Dewey did, Popper simply ignores the principle of synechism and breaks the solution of continuity that exists between the world of science and the life-world.[43]

To properly conceive of the natural sciences, Popper simply won't do. Like all forms of naturalism, scientism lacks reflexivity. Reducing the world to a billiard table without players to move the balls (an empty Fiasco bar, as it were…), it cannot even conceptualize its own practices. Popperianism is a "nocturnal" philosophy of science, as Bachelard would say. It is a philosophy for scientists, useful for getting grants perhaps, but not a philosophy of scientists. As soon as scientists start to experiment, the theory of falsification is falsified. As soon as they start talking, it is overcome.
One way or another, a connection with the human sciences has to be established. What we are looking for is not causes that determine the actions from without, but the symbolic meanings that orient actors from within. What we need to make sense of the social world are reasons rather than causes. Popper arrived at a similar conclusion, though he did not openly say so. Contra widespread opinion — shared by all my colleagues in the human sciences who put Popper in the curriculum — our Viennese philosopher is proposing different objectives and methodologies for the natural and the social

[40] But with Feyerabend we might ask: What does Popper actually know about astrology?

[41] Bhaskar, R.: *A Realist Theory of Science.*

[42] Apel, K.O.: *Die Erklären-Verstehen Kontroverse in transzendental-pragmatischer Sicht* and Habermas, J. *Erkenntnis und Interesse*.

[43] Cf. Haack, S.: "Not cynicism, but ***synechism***: Lessons from classical pragmatism", *Transactions of the Charles S. Peirce Society* 41:22, 239-253.

sciences.[44] The sciences may be loosely unified by the method of falsification, they are certainly not unified by the method of causal explanation. In the social sciences, we need to understand the motives of the actors if we are to explain their behaviour rationally. This is no doubt the right moment to invoke the *genius loci* of this place: Max Weber. So let me quote the opening paragraph of *Wirtschaft und Gesellschaft*: "Soziologie soll heißen: eine Wissenschaft, welche soziales Handeln deutend verstehen und dadurch in seinem Ablauf und seinen Wirkungen ursächlich erklären will". Following Max Weber, Popper affirms that in the social sciences, one has to work towards a rational explanation of action. He expressly states that rational explanations are not causal explanations, and wasted a good deal of his energy battling against historicist and futurologist invocations of the laws of history. After all, what distinguishes the human world from the social world is that the latter is a product of Man, whereas the former, allegedly, was created by God.

Unlike Weber, however, who followed Vico's humanist principle of interpretation – the *verum factum* principle according to which we can understand what we have made, but have to explain what we cannot understand – Popper is not really interested in hermeneutics and phenomenology. Although the reference to the method of explanatory understanding suggests a qualitative-interpretative approach of the social world in terms of shared cultural meanings actors actualize in their everyday behaviour, Popper has recourse to the modeling of marginal economics. Instead of a cultural reinterpretation of economic action, we thus get an economic interpretation of meaningful action!

To explain the predominance of instrumental action in modern societies, Weber proposed a cultural sociology of Western rationalism. Following Hayek, who hired him at the LSE, Popper reduces action to rational choice and eliminates culture altogether. Instead of offering a second order interpretation of the meanings that actors give to their action – a second order interpretation that reconstructs the reasons that motivate actors to pursue certain ends with reference to ultimate values, norms and beliefs – Popper takes the ends as given. Assuming with Weber that ends are ultimately arbitrary and subjective, he limits understanding to the rational calculation of the means that a hypothetical actor would have made if he were fully informed about the situation. By introducing a hypothetical actor who would act as Hayek (or any other rational fool) would, Popper has, however, severed the living link of culture that connects the social sciences to the humanities.

From this point of view, his invocation of methodological individualism (which is hardly compatible with his later theory of "world 3") can better be understood as an attempt to strip the social sciences of all meaning. When references to meanings, values, norms and beliefs are dispensed with as idealist waffle, material interests are all that remain. But if that is the case, then freedom necessarily disappears as well.[45] The protestations to the contrary of methodological individualism should not be taken at face value. If action is reduced to rational choice and the material conditions of action are known, the rational course of action can be determined almost automatically. The actor is reduced to a mere automaton that has no other choice but to act in the most rational fashion if he is to accomplish his ends. All one needs are some logarithms and the optimal course of action can be calculated with precision.

Of course, one could object that the rational man is only a methodological fiction – Weber would say an idealtype – and that it functions better the more real action deviates from the model. But then the question of falsification comes back with a vengeance. By launching the social sciences in the orbit of marginal economics, Popper has transformed the social sciences into analytic sciences that are true, come what may, by definition. Instead of an explanation of the course of action in real life that systematically interrelates ends and means with the values, norms and beliefs the subjects actually adhere to, Popper has nothing else to offer but a series of tautologies that cannot be refuted by reality. Once again, we can see that the theorist of falsification has immunized his own theory against the falsification by the experience and common sense of real people.

And now to finish my talk on a more positive note, I'd like to suggest a humanist alternative to the utilitarian calculus of preferences: Instead of rational choice, we need to pay more attention to internal voice and listen to what people actually say, what moves them and what they really care

[44] Cf. Werlen, B.: *Society, Action and Space. An Alternative Human Geography.*

[45] Cf. Parsons, T.: *The Structure of Social Action*.

about. If only we would listen more carefully to the internal conversations people have with themselves concerning what they want to do not so much in as with their lives, we will find that people are not only driven by material interests, but that they are also moved by ideals, principles and values.[46] Precisely because we are living in dark times, we desperately need some glimmers of hope.

[46] Cf. Vandenberghe, F.: "Language, Self and Society. Hermeneutic Reflections on the Internal Conversations That We Are", forthcoming in Archer, M. (ed.): *Conversations on Reflexivity*.


European
University
Institute
European University Press