

# Philosophy of the Social Sciences

## EDITORIAL BOARD

### *Managing Editor*

Ian C. Jarvie, *York University, Canada*

### *Editors*

J.N. Hattiangadi, *York University, Canada*

John O'Neill, *York University, Canada*

### *Editorial Board*

Joseph Agassi, *York University, Canada*

Hans Albert, *University of Mannheim, Germany*

David Bakan, *York University, Canada*

L.A. Boland, *Simon Fraser University, Canada*

Augustine Brannigan, *University of Calgary, Canada*

James M. Buchanan, *George Mason University, United States*

K.O.L. Burrige, *University of British Columbia, Canada*

Noam Chomsky, *Massachusetts Institute of Technology, United States*

Avi J. Cohen, *York University, Canada*

Kurt Danziger, *York University, Canada*

Jacques Derrida, *École des Hautes Études en Sciences Sociales, France,  
and Yale University, United States*

W.H. Dray, *University of Ottawa, Canada*

Morris Eagle, *York University, Canada*

Jon Elster, *University of Chicago, United States*

J.W. Grove, *Queen's University, Canada*

Irene Harvey, *Pennsylvania State University, United States*

Karin Knorr-Cetina, *University of Bielefeld, Germany*

Noretta Koertge, *Indiana University, United States*

Steven Lukes, *European University Institute, Italy*

Robert K. Merton, *Columbia University, United States*

Peter Munz, *Victoria University of Wellington, New Zealand*

Christopher Nichols, *York University, Canada*

Michael A. Overington, *St. Mary's University, Canada*

Alexander Rosenberg, *University of Georgia, Atlanta, United States*

Paul A. Roth, *University of Missouri–St. Louis, United States*

Marshall Sahlins, *University of Chicago, United States*

Dorothy E. Smith, *OISE, Canada*

Charles Taylor, *University of Montreal, Canada*

Sari Thomas, *Temple University, Philadelphia, United States*

H.T. Wilson, *Administrative Studies, York University, Canada*

J.W. Yolton, *Rutgers University, United States*

For Sage Publications: Mark Gage, Nicole Hirschman, Skye Loyd, and Joe Cribben

---

# Philosophy of the Social Sciences

---

Volume 31, Number 4

December 2001

## Articles

### **Between Social Science and Social Technology:**

*Toward a Philosophical Foundation for  
Post-Communist Transformation Studies*

ANDREAS PICKEL

459

### **Explanatory Unification:**

*Double and Doubtful*

USKALI MÄKI

488

### **"Agency" as a Red Herring in Social Theory**

STEVEN LOYAL

BARRY BARNES

507

## Discussions

### **Sexual Harassment and Wrongful Communication**

EDMUND WALL

525

### **Rewriting Color**

B.A.C. SAUNDERS

J. VAN BRAKEL

538

## Review Essay

### **Freeman on Mead Again**

I. C. JARVIE

557

## Book Review

*Explaining Culture: A Naturalistic Approach,*  
by Dan Sperber

MAHESH ANANTH

563

### **Index**

572

### **Erratum**

576

**PHILOSOPHY OF THE SOCIAL SCIENCES** is an international quarterly which publishes articles, discussions, symposia, literature surveys, translations, and review symposia of interest both to philosophers concerned with the social sciences and to social scientists concerned with the philosophical foundations of their subjects. No school, party, or style of philosophy of the social sciences is favored. Debate between schools is encouraged.

**MANUSCRIPTS** should be submitted in triplicate to Ian C. Jarvie, Managing Editor, Department of Philosophy, York University, North York, Ontario, Canada M3J 1P3. A guide for the preparation of manuscripts is available on request.

**BOOKS FOR REVIEW** should be sent to the editorial office above.

---

**PHILOSOPHY OF THE SOCIAL SCIENCES** (ISSN 0048-3931) is published four times annually—in March, June, September, and December—by Sage Publications, 2455 Teller Road, Thousand Oaks, CA 91320. Copyright © 2001 by Sage Publications. All rights reserved. No portion of the contents may be reproduced in any form without written permission of the copyright owner.

**Subscriptions/Inquiries:** Annual subscription rates for institutions and individuals are based on the current frequency. Prices quoted are in U.S. dollars and are subject to change without notice. Canadian subscribers add 7% GST (and HST as appropriate). Outside U.S. subscription rates include shipping via air-speeded delivery. Institutions: \$285 (within the U.S.) / \$301 (outside the U.S.) / single issue: \$83 (worldwide); Individuals: \$73 (within the U.S.) / \$89 (outside the U.S.) / single issue: \$29 (worldwide). Noninstitutional orders must be paid by personal check, VISA, or MasterCard.

Address all correspondence to Sage Publications, 2455 Teller Road, Thousand Oaks, California 91320. All subscription inquiries, orders, and renewals with ship-to addresses in North America, South America, Australia, China, Indonesia, Japan, Korea, New Zealand, and the Philippines must be addressed to Sage Publications, 2455 Teller Road, Thousand Oaks, CA 91320, U.S.A., telephone (800) 818-SAGE (7243) and (805) 499-9774, fax (805) 375-1700. All subscription inquiries, orders, and renewals with ship-to addresses in the U.K., Europe, the Middle East, and Africa must be addressed to Sage Publications, Ltd, 6 Bonhill Street, London EC2A 4PU, England, telephone +44 (0)20 7374 0645, fax +44 (0)20 7374 8741. All subscription inquiries, orders, and renewals with ship-to addresses in India and South Asia must be addressed to Sage Publications Private Ltd, P.O. Box 4215, New Delhi 110 048, India, telephone (91-11) 641-9884, fax (91-11) 647-2426. Address all permissions requests to the Thousand Oaks office.

Periodicals postage paid at Thousand Oaks, California, and at additional mailing offices.

This journal is abstracted or indexed in **Academic Search, Arts & Humanities Citation Index, Corporate ResourceNET, Current Citations Express, Current Contents: Arts & Humanities, International Bibliography of the Social Sciences, International Political Science Abstracts, ISI Basic Social Sciences Index, LEXIS-NEXIS, MasterFILE FullTEXT, Periodical Abstracts, Political Science Abstracts, Social Science Source, Social Sciences Citation Index, Social Sciences Index Full Text, Sociological Abstracts, SRM Database of Social Research Methodology, Standard Periodical Directory (SPD), The Philosopher's Index, TOPICsearch, Wilson OmniFile V, and Wilson Social Sciences Index/Abstracts**, and is available on microfilm from University Microfilms, Ann Arbor, Michigan.

**Back Issues:** Information about availability and prices of back issues may be obtained from the publisher's order department (address below). Single-issue orders for 5 or more copies will receive a special adoption discount. Contact the order department for details. Write to the London office for sterling prices.

Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by Sage Publications, for libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that the base fee of 50¢ per copy, plus 10¢ per copy page, is paid directly to CCC, 21 Congress St., Salem, MA 01970. 0048-3931/2001 \$.50 + .10.

**Advertising:** Current rates and specifications may be obtained by writing to the Advertising Manager at the Thousand Oaks office (address above).

**Claims:** Claims for undelivered copies must be made no later than six months following month of publication. The publisher will supply missing copies when losses have been sustained in transit and when the reserve stock will permit.

**Change of Address:** Six weeks' advance notice must be given when notifying of change of address. Please send old address label along with the new address to ensure proper identification. Please specify name of journal. POSTMASTER: Send address changes to: Philosophy of the Social Sciences, c/o 2455 Teller Road, Thousand Oaks, CA 91320.

# Between Social Science and Social Technology

## Toward a Philosophical Foundation for Post-Communist Transformation Studies

ANDREAS PICKEL  
*Trent University*

This analysis examines fundamental questions at the intersection of social science and social technology as well as problems of disciplinary divisions and the challenge of cross-disciplinary cooperation. Its theoretical-empirical context is provided by post-communist transformations, a set of profound societal changes in which institutional design plays a central role. The article critically reappraises the contribution of Karl Popper's philosophy to this problem context, examines neoliberalism as social science and social technology, and examines the role of experts and disciplinary divisions in the reform process. Building on Mario Bunge's social philosophy, it sketches basic elements of a cross-disciplinary approach to "social change by design."

Today it has become fashionable in the sciences to appeal to the specialized knowledge and authority of experts, and fashionable in philosophy to denigrate science and rationality. Oftentimes, this denigration of science and rationality is due to a mistaken *theory* of science and rationality—a theory which speaks of science and rationality in terms of specializations, experts, and authority. But science and rationality have really very little to do with specialization and the appeal to expert authority. On the contrary, these intellectual fashions are actually an obstacle to both. For just as the fashionable thinker is a prisoner of his fashion, the expert is a prisoner of his specialization. And it is the freedom from intellectual fashions and specializations that makes science and rationality possible. (Popper 1994, ix)

---

*Received* 25 November 1999

The author wishes to thank Hans Albert, Dariusz Aleksandrowicz, Mario Bunge, David Clarke, and two anonymous referees for useful comments and suggestions. Funding from the Social Science and Humanities Research Council of Canada is gratefully acknowledged.

*Philosophy of the Social Sciences*, Vol. 31 No. 4, December 2001 459-487  
© 2001 Sage Publications

## CONTEXT

The present analysis examines fundamental questions at the intersection of social science and social technology as well as problems of disciplinary divisions and the challenge of cross-disciplinary cooperation. What follows is primarily a philosophical analysis that has as its substantive theoretical and empirical focus the general field of post-communist transformation studies. Emerging in the aftermath of the communist regime collapse in Eastern Europe and the Soviet Union, this is a relatively young field that has attracted scholars from a wide range of disciplines—economics, political science, sociology, geography, demography, and psychology, as well as history, cultural studies, and philosophy. The field was quickly acknowledged as one of the best “social science laboratories” (Offe 1991) for disciplines that rarely have the opportunity to undertake their empirical studies under “laboratory” conditions. However, it is not just the great speed, the profound and comprehensive character, and the contemporaneity of social changes in the post-communist regions that qualify them as quasi laboratories. In the age of globalization, few places are spared the powerful effects of accelerated social change. Yet nowhere have we witnessed such conscious and concerted attempts to steer and control these macroprocesses of social change toward a well-defined set of systemic goals—attempts usually referred to as “making the transition to the market and democracy.”<sup>1</sup> The transition doctrines, plans, and policies that have been produced, reached dominance, and have been implemented to varying degrees, constitute what at any rate could be considered the most advanced forms of large-scale social technology at the end of the 20th century, as I will further explain below. As such, it is the design element of these change processes that makes them particularly relevant for social studies in general. In other words, what enhances the quasi-laboratory conditions of the post-communist regions is the fact that experiments in social engineering (to use Popper’s phrase)<sup>2</sup> are in fact being carried out, if not under the direct guidance of scientists themselves, then by politicians with the help of experts who presumably draw on scientific knowledge. Post-communist studies thus is an excellent field for studying the relationship between social science and social technology and the challenge of cross-disciplinary cooperation. In the spirit of critical rationalism, this article will revisit some of Karl Popper’s ideas of relevance in this context and try to improve upon them.

## POPPER AND TRANSFORMATION STUDIES

My *social theory* (which favours gradual and piecemeal reform, reform controlled by a critical comparison between expected and achieved results) contrasts with my *theory of method*, which happens to be a theory of scientific and intellectual revolution. (Popper 1994, 68)

The most relevant contributions of Popper's philosophy to the field of transformation studies I take to be (1) the problem-oriented approach and (2) the critique of utopian social engineering. The first is particularly significant for questions of disciplinary division and cooperation. The second relates to the role of social science and social technology in social change. I will briefly introduce my own adaptation of the problem-oriented approach to transformation studies in the following section. Suffice it to say at this point that Popper's insistence on the primacy of problems, both theoretical and practical, in advancing scientific knowledge relegates disciplinary boundaries and frameworks to the status of administrative divisions—whose intrinsic theoretical merit, if any, is more than offset by their tendency to stifle scientific inquiry (see, e.g., Popper 1970, 1994).

Popper's critique of utopian social engineering (Popper 1945, 1957), developed in the context of the communist experiment of revolutionary social change by design, as well as his endorsement of "piecemeal social engineering," has remained highly relevant for the post-communist experiments (Dahrendorf 1990; Isaac 1996; Kabele and Radzai 1993; Murrell 1992, 1995; Soros 1998; Sullivan 1994). Ralf Dahrendorf, for example, in his *Reflections on the Revolution in Europe* (first published in 1990), explicitly refers to Popper's conception of utopian versus piecemeal social engineering as he reflects on the tasks facing reformers in Eastern Europe immediately after the collapse of communism. Dahrendorf himself does not believe that Popper's conception is helpful for the problem at hand. "Even apart from the unfortunate connotations of social engineering, 'piecemeal' is not quite enough when one is faced with a constitutional challenge" (Dahrendorf 1990, 161). In a decade of work in the field of post-communist studies, I have gradually come to the conclusion that Dahrendorf is right: Popper's conception is ultimately inadequate and even misleading, as I will try to show below. However, this is not to say that it does not contain valuable and highly relevant insights for some fundamental problems of social science and social technology. My own sense in 1989-90, while just completing a dissertation that used some

of Popper's arguments and insights to account for the peculiar and unintended consequences of East German private sector social engineering under communist rule (Pickel 1992), was that the radical blueprints for the rapid transition to capitalism known as economic shock therapy could fruitfully be approached as contemporary instances of utopian social engineering. The proponents of radical, comprehensive reform doctrines as well as many other participants and observers of the events, by contrast, assumed that whereas the goal of socialism had been utopian, establishing a liberal market order supposedly was not. As a result, unlike his conception of the open society, Popper's critique of utopian social engineering was not really considered applicable to post-communist Eastern Europe. Proponents of radical reforms feared primarily the political and bureaucratic opposition of antireformist interests rather than other, deeper obstacles in the way of reaching the nonutopian goal of the market. But it was precisely some of those deeper obstacles that Popper had in mind when he rejected "wholesale social engineering" as utopian. True, the context for his original critique was the Marxist-Leninist project of social transformation. But the force of his argument was directed against a specific, *holistic approach* to social reform, not simply against this or that set of utopian goals. Thus, his critique applies equally to the neoliberal project of radical marketization. In Popper's words, "Holistic or Utopian social engineering, as opposed to piecemeal social engineering . . . aims at remodeling the 'whole of society' in accordance with a definite plan or blueprint."

The piecemeal engineer knows, like Socrates, how little he knows. He knows that we can learn only from our mistakes. Accordingly, he will make his way, step by step, carefully comparing the results expected with the results achieved, and always on the look-out for the unavoidable unwanted consequences of any reform; and he will avoid undertaking reforms of a complexity and scope which make it impossible for him to disentangle causes and effects, and to know what he is really doing. (Popper 1945, 67)

Popper's epistemological argument seems to condemn any comprehensive and fundamental social reform project, including post-communist marketization and democratization, as utopian, for the systemic changes are necessarily of a complexity and scope that transcend the careful, step-by-step method he favors. However, the crux of his argument against the holistic approach is the claim that such

wholesale changes turn out to be impossible to accomplish in practice. "The greater the holistic changes attempted, the greater are their unintended and largely unexpected repercussions, forcing upon the holistic engineer the expedient of piecemeal *improvization*."

It continually leads the Utopian engineer to do things which he did not intend to do; that is to say, it leads to the notorious phenomenon of *unplanned planning*. Thus the difference between Utopian and piecemeal social engineering turns out, in practice, to be a difference not so much in scale and scope as in caution and preparedness for unavoidable surprises. One could also say that, in practice, the two *methods* differ in other ways than in scale and scope—in opposition to what we are led to expect if we compare the two *doctrines* concerning the proper methods of rational social reform. (Popper 1945, 68-69)

This fundamental insight, derived from Popper's fallibilist epistemology, contributes significantly to our understanding of why radical liberalization programs in so many former communist countries were never fully implemented (Pickel and Wiesenthal 1997). They were conceived as a revolutionary process of "planned unplanning" and have ended up, precisely as Popper predicts, as processes of "unplanned planning." Yet this central insight alone does not help us to explain the very significant differences in levels of success and failure attained by different reform countries in their clearly "holistic"—that is, comprehensive and large-scale, systemic—reform projects. The reason is that Popper's epistemological insight is only one, albeit important, part of the story. It concerns directly only the limitations of our social scientific knowledge. That epistemological fallibilism has implications for social technology should be evident. However, what these implications are, and more specifically, what the relationship between social science and social technology is, could be, and should be, is barely examined in Popper's work (see, however, Agassi 1985; Albert 1976; Bunge 1998; Dryzek 1990; Fischer 1993). Radical reformist doctrines of comprehensive liberalization ("planned unplanning") were not primarily scientific claims (although all too often they were claimed to have scientific status). Rather, they were above all action plans for key reform policies and political platforms in the struggle for political power and ideological hegemony. As such, should they really be subject to the same criticisms as scientific theories?



THE FUNDAMENTAL PROBLEMS  
OF POST-COMMUNIST TRANSFORMATION

The work of the scientist does not start with the collection of data, but with the sensitive selection of a promising *problem*—a problem that is significant within the current *problem situation*, which in its turn is entirely dominated by our theories. . . . Scientific problems are preceded, of course, by pre-scientific problems, and especially by practical problems. (Popper 1994, 155-56)

In my own work on post-communist transformations, I have tried to explore the relationship between science, policy, and politics in somewhat greater depth. In this I was inspired by Popper's problem-oriented approach, his second major contribution of relevance to transformation studies. While Popper has nowhere systematically developed such a problem-oriented approach, it seems nevertheless fundamental to his fallibilist philosophy. In short, knowledge without absolute foundations that nonetheless is claimed to advance through trial and error requires nonfoundational criteria for identifying and criticizing our mistakes. Popper's disciple W. W. Bartley has proposed four methods of criticism that can be applied to empirical theories, metaphysical assumptions, or political norms without themselves being based on any epistemological authority.

We have at least four means of eliminating error by criticizing conjectures or speculations. These checks are listed in descending order according to their importance and the rigor with which they can be applied. (1) The check of *logic*: Is the theory in question consistent? (2) The check of *sense observation*: Is the theory empirically refutable by some sense observation? And if it is, do we know of any refutations of it? (3) The check of *scientific theory*: Is the theory, whether or not it is in conflict with sense observation, in conflict with any scientific hypotheses? (4) The check of the *problem*: What problem is the theory intended to solve? Does it do so successfully?" (Bartley 1984, 127)

In the following quotation, Popper himself elucidates what is implied by the check of the problem.

Every *rational* theory, no matter whether scientific or philosophical, is rational in so far as it tries to *solve certain problems*. A theory is comprehensible and reasonable only in its relation to a given *problem-situation*, and it can be rationally discussed only by discussing this relation. Now if we look upon a theory as a proposed solution to a set of problems,

then the theory immediately lends itself to critical discussion—even if it is non-empirical and irrefutable. For we can now ask questions such as, Does it solve the problem? Does it solve it better than other theories? Has it perhaps merely shifted the problem? Is the solution simple? Is it fruitful? Does it perhaps contradict other philosophical theories needed for solving other problems? (Popper 1963, 199)

In my attempt to come to grips with the status and function of reformist doctrines in the post-communist context, it became increasingly clear to me that their significance and power in the debate had little to do with their scientific credentials. Indeed, radical reform advocates have typically combined mainstream neoclassical economics, historicism, and market essentialism into an impressive looking “scientific” foundation for systemic transformation that serious social scientists could not possibly accept as such (Pickel 1995, 1997). Yet, radical market reformism has remained politically dominant, and it has arguably produced both transformation successes and failures. How could a scientifically weak doctrine remain so strong politically and even be implicated in some policy successes? To answer this somewhat puzzling question, let us begin by broadening Popper’s problem-oriented approach.

Post-communist transformation processes relate to a host of fundamental theoretical problems in the social sciences—probably the main reason why it has become such an attractive field of study for scholars in these disciplines. Reform doctrines can be examined in terms of their theoretical content: any contemporary social reform doctrine that is internally inconsistent, empirically untenable, or seriously at odds with current social science knowledge is unlikely to be translated into rational, science-based reform policy (i.e., social technology). Neoliberalism, as we will see in a moment, is an instance of such a pseudoscientific, or at best protoscientific, reform doctrine. Problems of post-communist transformation *at the theoretical level* are problems of conceptualizing and explaining the changes under way in formerly or still communist countries in the context of a larger body of knowledge concerning processes of fundamental social change. Of particular importance at this level is the problem of fundamental reform, that is, the conditions for the possibility of controlled social change. The scientific evidence concerning the possibility of controlled holistic reform programs has in fact led most theorists, quite in line with Popper, to conclude that in modern complex societies such projects of planned systemic change are practically impossible

(Wiesenthal 1997). However, while reform doctrines can and should be assessed by such scientific standards, this alone would give us only a very partial view of their overall strength. The reason is that reform doctrines are designed as well, or perhaps above all, to solve *different types of problems*.

With the sudden and unexpected collapse of communist regimes in 1989, there was an immediate *practical*, that is, strategic and policy, need for suitable reform doctrines. At the time, a political consensus quickly emerged that a controlled systemic transition from a centrally planned economy to a capitalist market economy was possible (Berend 1999; Szacki 1996), so the debate shifted largely to questions of strategy and policy. The problem context from this strategic, policy point of view differs in fundamental respects from the social scientific problem context. The strategic debate largely ignored the social sciences' "impossibility theorem" and for its practical purposes, adopted the opposite working assumption: the systemic transition from communism to capitalism can be achieved through a radical institutional break with the past and the rapid introduction of a "market infrastructure." Like any social technology designed to effect change, post-communist reform doctrines identify a given state of affairs as unsatisfactory, spell out the alternative state of affairs or goal to be reached, and assign the appropriate means or policies. Unlike social science, social technology is based on political norms and moral standards that determine what constitutes an unsatisfactory state of affairs, a desirable goal, and acceptable means. The particular problem situation after the collapse of communism favored rather stark and simplistic answers to these basic questions: the communist system was bankrupt, capitalism and liberal democracy were desirable and indeed "natural" or at least "historically inevitable," and systemic change required little more than the elimination of the old system plus a set of liberalization measures. This radical simplification of reality and drastic conceptual reduction of the problem to a few crucial variables responded powerfully to the fundamental problem of strategy and policy in 1989. Richer, more nuanced conceptions (cf. gradualism; Murrell 1992; Poznanski 1995) proved to be too complex and normatively too ambiguous to provide competitive answers to the question What is to be done? Of course, a social technology should not be judged solely in terms of its prospects for being adopted. The most basic standard to apply to any social technology is Does it work? Yet, a social technology that is never adopted by definition cannot

pass the test. A social technology that works barely, or generates mixed success, on the other hand, can be defended by its proponents in a variety of ways (Aslund 1995). This takes us from the level of strategy to the level of ideology and politics.

From the definition of the political agenda to the legitimation of the political order, there is a range of problems that any society has to cope with to create a minimum of social stability, internal peace, and external security, and thus the preconditions for the possibility of achieving more ambitious societal goals, such as stable democracy, prosperity, and regional integration. Societies undergoing rapid and profound changes experience these problems much more acutely than stable societies where established institutional mechanisms and routines may be so successful that these fundamental problems seem nonexistent. Ideologies provide responses to these fundamental problems and are crucial in developing and maintaining these more permanent institutional mechanisms and routines. In this sense, the economic transformation of a society is perhaps above all a political challenge. The challenge of economic reform poses a variety of specific ideological and political problems: forging a vision for the new order, mobilizing voters for the reform program, reaching elite consensus on policies, defining and securing the country's place in the larger world, providing legitimation of policy outcomes and the emerging new order, maintaining the credibility of and commitment to the reform project, and holding political opponents in check. Any social technology for fundamental change has to prove itself in such an ideological and political problem context. Its relative success in those respects will determine not only whether a reform doctrine will be adopted but also whether it will have a social and political environment within which, once adopted, it can work.

In sum, social technologies such as post-communist reform doctrines should be evaluated using a problem-oriented approach. However, it is crucial to take into account the diverse problems and problem contexts in which social technologies have to prove themselves. We have discussed three general sets of problems and contexts: scientific knowledge (Is the doctrine true?), strategy and policy (Does it work?), and ideology and politics (Does it sell and do its clients remain faithful?). Let us now illustrate our problem-oriented method by looking more closely at the most successful post-communist reform doctrine—neoliberalism.

## NEOLIBERALISM AS SOCIAL TECHNOLOGY

Neoliberalism is probably the most influential doctrine of social change in the late 20th century.<sup>3</sup> In the light of our distinction between three fundamental types of transformation problems—theoretical, strategic, and political—let us now look more closely at neoliberalism. As with any theory, doctrine, or ideology, attempts to capture its “essence” can always be criticized as inaccurate, misleading, or simply a convenient target constructed by the critic for subsequent demolition. Our purpose here is not simply to expose neoliberalism as ideology. To acknowledge the fact that it provides answers to fundamental political and ideological questions is therefore not a criticism as such. There are no nonideological answers to these questions.<sup>4</sup> From a scientific point of view, we would ask the theoretical and empirical question whether and to what extent the ideology is politically successful in a given problem context. From a normative point of view, we would ask moral and political questions about the ideology’s defensibility (even if successful, it may be reprehensible; cf. Standing 1998).

The political success of neoliberalism as an ideology of post-communist change in Eastern Europe is largely due to its contextual fit; that is, to the way in which it was able to address the fundamental problems of order and legitimacy after the collapse of communism and the specific ideological and political problems of systemic reform sketched out above. At the same time, it is important to note the limits of neoliberalism’s success as an ideology in Eastern Europe, especially for purposes of domestic agenda setting, mobilization, and legitimization. Antiliberal, ethnic nationalism has been and continues to be a powerful force in the whole region. The further east and south we move, the less successful has been the ideological strength of neoliberalism in terms of its domestic political purchase. As the ideology and discourse of globalization and regional integration, its popularity and situational fit seem to reflect closely a particular country’s prospects and its population’s self-perceptions along these lines.<sup>5</sup>

But neoliberal reform doctrines have also been extremely influential as social technologies for economic reform. In fact, most proponents of neoliberalism would flatly deny my characterization of their reform doctrines as ideological, a charge they usually reserve for their opponents. Neoliberals routinely invoke neoclassical economics as the scientific basis of their social technologies of market transition. We

will examine the quality and strength of this basis below. Let us first note, however, that social technologies cannot be directly derived from scientific knowledge because they are neither politically nor morally neutral. (Conversely, reform doctrines that derive mainly from ideological precepts with little or no basis in current scientific knowledge should not be considered social technologies but merely techniques.) Neoliberals tend to downplay or ignore the unavoidable normative dimensions of reform technology. In fact, by playing on the scientific credentials of economics as a discipline, neoliberals have been successful in selling their doctrines as expert social technology and progressive ideology to the powerful and even the less powerful.<sup>6</sup> The considerable sophistication with which neoliberal transition policies have been sold to reformers and many of the to-be-reformed suggests that this dimension of promoting their doctrines might itself be considered a social technology—more specifically as the cultural and political technology of spreading economic ideas (on Keynesianism, see Hall 1989; on neoliberalism, see George 1999).

There continues to be a heated debate on the relative success of neoliberal reforms in post-communist Eastern Europe. Admittedly, the apparent simplicity of the question whether neoliberal social technology works, can work, or has worked is deceiving. Even if we accept the simplifying assumption that neoliberal reform doctrines were more or less followed for at least some time in the entire region (Zecchini 1997), the picture remains sufficiently ambivalent to support various, mutually inconsistent interpretations. They range from the dogmatic technocratic view that reform failures simply mirror inadequate adoption and implementation of radical reforms, to more nuanced views that take into account different initial conditions (Balcerowicz 1995), historical legacies (Jowitt 1992), social networks (Grabher and Stark 1997), or international dynamics (Bönker 1994; Chilton 1995) in the reform process. To complicate matters further, neoliberal reforms qua social technology can only be adequately judged if both empirical-theoretical and normative-political standards are applied. The debate tends to be hopelessly confused on this distinction, presenting normative (often patently ideological) judgments as scientific assessments, or conversely rejecting empirical evidence or theoretical arguments as ideological. Taking reform doctrines such as neoliberalism seriously as social technology, as is proposed here, injects some analytical clarity into this debate. Such doctrines are not just science and should therefore not be presented as

such, as many proponents of reform plans tend to do. At the same time, they are not—or at least may not be—just ideology, as their critics tend to charge (Gowan 1995). A social technology needs both a sound scientific basis and a sound moral-political basis. The remainder of the article will focus on the scientific basis of neoliberal social technology. This will give us an opportunity to attend more closely to questions of interdisciplinarity, our second major theme.

#### TRANSFORMATION STUDIES, EXPERTS, AND THE DISCIPLINARY DIVISION OF LABOR

A necessary condition for unified policy-making is a unified social science. (Bunge 1998, 452)

Since the beginning of the transformations in 1989, experts have played crucial roles in the fundamental restructuring of formerly communist economies and states. From the first wave of economic experts counseling shock therapy to the current wave of specialists in European Union (EU) rules and regulations (Jacoby 2002), a host of social technologies, mostly of Western origin,<sup>7</sup> have been adopted and adapted in the reform process. They include macroeconomic policies such as stabilization and liberalization, institutional reforms such as privatization and social sector restructuring, constitutional reforms of the state such as democratization and regionalization, as well as the introduction of innumerable new rules and regulations designed to create the framework of a liberal market economy and polyarchy on the remains of the communist system. There can be no doubt that the systemic switch from an old and exhausted model of industrialism and centralism to a new and dynamic informational economy and pluralistic politics depends for its success to a large extent on the expert knowledge of many different specialists. Whether it is banking reform, pension reform, electoral law, corporate law, or environmental regulation, only specialists are in a position to design the requisite social technologies to bring into being such political and economic institutions.

Many of the experts, and even more of the expert knowledge, have been imported from the West. This should be no surprise, since the economic and political systems to be created are those of Western capitalism and liberal democracy. What was missing, and what the West could not supply, was expertise in guiding such large-scale, complex,



and interrelated transformation processes. The available specialized knowledge was helpful to the extent that limited sociotechnical systems were to be implanted. It was unhelpful and often dangerous to the extent that such limited sociotechnical systems depended for their anticipated functioning on a variety of formal-legal, informal cultural, and other preconditions that were not in place and could not easily be transplanted from their original context. Many social technologies for reform were more or less arbitrary abstractions from specialized knowledge that was insufficiently aware of its own embeddedness in a particular economic, political, and cultural system. Thus, the utopian character of such reform policies was not due to excessive scope but rather to an excessively limited, sectoral conception of reality and a corresponding piecemeal approach to changing it that is insufficiently holistic in perspective. Traditional disciplinary divisions in the social sciences have played a major role in this. Mainstream neoclassical economics, the scientific basis for neoliberal social technology, is a case in point.

Economists and economic experts as a group probably enjoy the greatest influence and respect among social scientists in the West. Their authority is based in part on a carefully maintained image of neoclassical economics as a genuine science (as opposed to such allegedly prescientific disciplines as sociology and political science). Politicians and the public at large, however, are rarely interested in the scientific discoveries of mainstream economists. They are interested in advice and forecasts that address practical problems. However, as we have seen above, such policy advice is at best part science and always part ideology. The public authority of mainstream economists is closely related to their collective ability to have consumers of their expertise associate their policy advice with science, objectivity, and rationality, whereas the policy advice of dissenters is portrayed as unscientific, partisan, and ideological. The same mechanism, enhanced by the initial attractiveness of all things Western and the helping hand of Western financial institutions and political elites, has been transplanted to Eastern Europe and placed economic experts in the role of "wholesale social engineers."

As a scientific basis for social technologies of post-communist transformation, however, neoclassical economics has surprisingly little to contribute. First, while the transition from a command economy to a liberal market economy obviously poses a range of economic problems, it also, and simultaneously, poses a range of institutional, political, and cultural problems. Mainstream economists usually do



not claim specialized knowledge in the latter types of problems. In fact, their own claim to scientific status depends on a high degree of theoretical formalization that rests on a narrow and sectoral conception of the economy. (In addition, it is also based on highly problematic assumptions about human behaviour and society [Bunge 1998, chap. 3].) While the relevance of this knowledge for explaining economic phenomena even in established market systems is disputed, it is ruefully inadequate as a basis for systemic change by design. Neoclassical economics has no theory of society but only an idealized and empirically and theoretically precarious model of a market economy. It therefore has no theory of social change but only a set of implicit normative implications according to which, *ceteris paribus*, the closer the economic system is to the stylized market economy, the more efficient the outcomes. It has a highly problematic basis for drawing explicit normative or policy conclusions in welfare economics, but this is a pseudoscientific attempt to avoid addressing normative and political questions explicitly (Albert 1978, chap. 5; Bunge 1998, chap. 10; Myrdal 1954). As a result, policy advice presumably based on, or even "derived from," neoclassical economics is in large measure ideological, while at the same time denying, or failing to make explicit, its normative assumptions. The central values embedded in neoliberal ideology, the most common ideology to accompany neoclassical theory, places the individual above society, endorses a vision of the individual as a *homo oeconomicus*, accepts and defends high degrees of social inequality, and, partly because it has so little sense and understanding of the social, opposes any form of collective political action, except when in defense of the market order.

To be sure, the academic discipline of mainstream economics cannot be held primarily responsible for the consequences of any social technologies designed in its name, though barely on its basis. Mainstream economics as a science has, especially in its more recent move toward ever-greater formalization and mathematization, not even shown particularly strong interest in so-called real-world problems. It is thus a great irony that post-communist economic transformations were so powerfully influenced by the advice of economic experts whose scientific knowledge had so little relevance for the sociotechnological problems of systemic change.

Another branch of social science that sometimes claims to look at the big picture and has had some, albeit much less, influence on

post-communist transformations, is political science. The intrusion of so-called transitologists, that is, specialists on democratization, into the field of post-communist transformation studies has sparked a lively debate on the role of generalizing versus contextually limited approaches to studying democratization (Bunce 1995; Karl and Schmitter 1995). Unlike their colleagues in the marketization branch of transformation studies, transitologists have played at best a modest role in advising reform governments on political reform. The equivalent of the economic shock therapists in the early stages of transformation were human rights lawyers, who, in Popperian spirit, advocated “negative constitutionalism—the notion that constitutions have a primarily negative purpose of preventing tyranny” (Holmes 1995; cf. also Pickel 1989). The narrowness of both politico-technological and economic-technological approaches, however, cannot be blamed entirely on the social engineers who might have inadvertently or willfully ignored a vast body of social science knowledge, only waiting to be translated into post-communist reform plans. The problem was and is also one of divided, compartmentalized, and often irrelevant social science knowledge. Most important, none of the social sciences was able to provide the kind of integrated, indeed holistic, approach and overarching perspective that the practical problems of transformation required (Müller 1995; the only partial exception, neomarxist approaches, were politically discredited). In the breach jumped self-appointed experts and ideologues whose reform prescriptions were based on political attractiveness, common sense, intuition, and abstractions of Western models. The result was piecemeal social engineering, albeit in a negative sense: large-scale reform policies were launched in some sectors of society, resulting in major unintended consequences in others, leading to “unplanned planning” and generating enormous social and economic costs.

#### A SOCIAL PHILOSOPHY FOR A UNIFIED SOCIAL SCIENCE

A well-rounded social philosophy must include a positive theory of society along with a positive moral philosophy—that is, one positing social goods, however debatable and changeable. Without such a global and positive social philosophy, no clear vision of an open society can emerge. And without such vision, people won’t be mobilized to build the new society. (Bunge 1996b, 552)

Popper's social philosophy contains many useful suggestions for the theory and practice of transformation.<sup>8</sup> In particular, his strong critique of utopian blueprints for reform has remained of some relevance here, as I have tried to show above. An even more important methodological tool, in my view, is Popper's problem-oriented approach, even if he did not sufficiently elaborate it for problems of social science theorizing, planning, and implementing social changes. In fact, as Mario Bunge reminds us,

Although he favored planned social reform, Popper never put forth any constructive proposals for it. Moreover, he did not examine in detail any of the social technologies, such as normative macroeconomics, city planning, social medicine, the law, or management science, all of which raise interesting ontological and epistemological problems—such as, for instance, the question of the very nature of plans as different from theories. (Moreover, he was not entirely clear about the distinction between social science and social technology.) . . . By writing off all ideologies—and moreover without analyzing in any detail the very concept of an ideology—Popper locked himself out of political science and political philosophy. (Bunge 1996b, 542-43)

The problem of controlling political rulers, perhaps the centerpiece of Popper's social philosophy, identifies one, albeit not the only or even the central, problem of politics. Two at least equally fundamental problems are the problem of political order and the problem of political legitimacy (Pickel 1989; Eidlin 1997). Moreover, post-communist transformation is not confined to changes in the political system. It includes simultaneous and interrelated changes in economic and cultural systems, embedded in regional and global dynamics, that raise fundamental problems of their own. It is here that the limits of Popper's social philosophy become particularly evident.

Popper has had nothing original, let alone constructive, to say about any social order, actual or desirable, beyond that it should be nontyrannical and should involve the protection of the destitute. . . . Popper's social philosophy lacks a theory about social order because he has neither a theory of society nor a positive moral philosophy. All Popper's social philosophy does is admonish us to replace the substantive traditional question "Who shall be the rulers?" with the procedural question "How can we tame them?" . . . In other words, Popper's conception and defense of liberty and democracy is limited to law and politics, and even then only to their mechanics. It warns us against despotism but does not help us redesign society to remove the sources of tyranny. Hence Popper's praise of social engineering, though sincere,

rings hollow: it enjoins us to plan without specifying any goals other than freedom. The result is a negative, spotty, superficial, formalist, and at places inconsistent social philosophy. It bears no comparison with the social philosophies of Khaldun, Machiavelli, Spinoza, Hobbes, Locke, Montesquieu, Rousseau, Mill, Marx, or even Paine, Kropotkin, or Laski. In my opinion it is also inferior to Popper's own contribution to the theory of knowledge, in particular his successful demolition of inductive logic and defense of epistemological realism. (Bunge 1996b, 550-51)

Even if Popper's social philosophy is admittedly "thin," in the sense that it has a narrow view of politics, little to say about economics and culture, and is unhelpful for the crucial distinction between problems of social science, social technology, and ideology, does it not provide at least an innovative methodological approach to social theorizing in the form of the "logic of the situation"? As a matter of historical record, the logic of the situation has not had any significant influence in the social sciences. Popper's version of methodological individualism is at odds with the very popular and influential rational choice model (Agassi 1987). And it is obviously opposed to all approaches in the social sciences that espouse one or the other version of methodological holism. A recent symposium reexamining Popper's model of situational analysis has yielded a mixed verdict on its usefulness and relevance (Matzner and Jarvie 1998). Popper's methodological individualism recapitulates an ontological individualism in the tradition of Mill, Weber, and neoclassical economics. This individualism—both ontological and methodological—has significant merits as a critical response to holism or collectivism, but it reaffirms a questionable alternative. As Bunge points out, individualism commits us to the view that there are no social entities with supraindividual features. This is highly problematic for two reasons.

One is that every human being is part of several social systems—such as families, business firms, schools, clubs, and informal social networks—so that his behavior is unintelligible without reference to them. Another reason is that every social system is characterized by emergent or systemic properties, such as social structure, viability, cohesion, history, progress, decline, and wealth distribution. . . . Whoever denies the existence of social systems is bound to either smuggle them in or invent surrogates for them. Popper was no exception. Indeed, to explain individual actions, Popper invokes institutions and "situations" (or "states of affair") as other individualists invoke "contexts" and "circumstances. . . . The entire "logic of the situation" resorts then to supraindividual items . . . Popper's social ontology may therefore be

characterized as individualistic rather than as consistently individualistic. (Bunge 1996b, 532-33; more on this hybrid in Bunge 1996a)

What alternative, substantive social philosophy is there for a critical rationalist working on problems of post-communist transformation—one that would be consistent with a problem-oriented approach, would empower us to transgress disciplinary borders, and give us tools to deal with the relationship between social science and social technology?

#### TOWARD A CROSS-DISCIPLINARY APPROACH TO POST-COMMUNIST TRANSFORMATION

Constructive action, whether individual or social, calls for positive views and plans in addition to rational discussions of goals and means. In particular, the design, planning, and construction of a better social order requires more than a handful of danger signals to help avoid or fight tyranny: it also calls for a positive social philosophy including a clear vision of the open society—one capable of motivating and mobilizing people. (The warning “Here be dragons” may be helpful, but it does not point to the right way.) And such a philosophy had better form a system rather than an aggregate of disjoint views, for social issues—like any correct ideas about them—happen to come in bundles, not one at a time. One step at a time, yes; one thing at a time, no. (Bunge 1996b, 553-54)

Much of the transition literature follows conventional disciplinary lines of inquiry. This is as true for the economics literature as for the political science literature on post-communist transformation, the two fields that account for the bulk of scholarly production in the first decade of post-communist “transitology.” Much of this literature, even when it has not explicitly taken the form of policy advice, appears to be normatively and ideologically driven. Specifically, economists are working on the premise that the telos of transformation is the establishment of a Western-style market economy, while political scientists are studying the conditions for and obstacles to democratization and democratic consolidation along Western lines. The problem is not the presence of normative and ideological assumptions as such but rather the often implicit adoption of the view that the radical change processes occurring in post-communist and other reforming countries are best conceptualized as “catching-up modernization” (Habermas) and Westernization. From this perspec-

tive, it appears reasonable to approach the study of transformation as a country's successful or unsuccessful approximation of an—often idealized—Western economic, political, and cultural model, each dimension of which is best studied on its own by the competent social science discipline.

Along with other critical voices in the debate (see, among others, Grabher and Stark 1997; Greskovits 1998; Rona-Tas 1997), I contend that the Westernization or convergence thesis is not the most appropriate, and a potentially quite misleading, point of departure for the formulation of the fundamental problems of transformation. It uncritically accepts the premises and goals of a political project as basic assumptions for social science theorizing. I propose, by contrast, to conceptualize systemic changes “on their own terms” and in a more open and broad fashion. More concretely, we should remain sensitive to the different degrees to which the political project of Westernization is pursued by various actors and as such constitutes important causal factors that need to be closely examined. However, we need to be equally interested in investigating the actual dynamics of change in the transforming societies that cannot be captured by viewing phenomena of social change discretely as more or less successful instances of economic, political, or cultural Westernization. In the remainder of the article, I sketch out a social theoretical foundation for transformation studies that is consistent with a problem-oriented approach, helps us to transgress disciplinary boundaries, and gives us tools to deal with the relationship between social science and social technology. Most of its central ideas are drawn from the philosophy of Mario Bunge.

#### Framework: Systemism

*Systemism* is an alternative to both individualism and holism. It accounts for both individual agency and social structure. It postulates that everything is a system or a component of one. It models every system in terms of composition, environment, and structure. It breaks down society into four major subsystems—biological or kinship system, economic system, political system, and cultural system. It can be applied at subnational, national, and transnational levels of analysis. Above the level of societies organized within nation-states, there are regional “supersocieties” (e.g., EU) as well as the world social system. It is important to note that, unlike Parsonian systems theory, systemism is not a theory but only a framework or approach, “just a

skeleton to be fleshed out with specific hypotheses and data" (Bunge 1996a, 265; further on systemism, Bunge 1996a, chap. 10; 1998, chap. 6.3). As such, it is not "static" or blind to processes of conflict and change.

Adoption of the systemic approach will avoid the pitfalls and tunnel vision which the narrow specialist invariably falls into, incapable as he is of taking into account any features that are not studied in his field. In other words, systemism favors interdisciplinarity and multidisciplinary. By the same token, it helps to avoid the costly mistakes made by the specialist—scientist or technologist, policymaker or manager—who overlooks most of the features of the real system he studies, designs, or steers. (Bunge 1996a, 266)

#### Key Theoretical Concept: Change Mechanisms

*Mechanismic* explanation is not to be confused with mechanical explanations. "Whereas a few of the mechanisms studied by contemporary science and technology are mechanical, most are not. Indeed, there are mechanisms of many kinds: electromagnetic, nuclear, chemical, cellular, intercellular, ecological, economic, political, and so on. Any explanation involving reference to a mechanism may be said to be *mechanismic*" (Bunge 1997, 411). Mechanismic explanation differs from most standard modes of explanation employed in the social sciences: the neopositivists' "covering law" model of scientific explanation, the interpretive approach of the hermeneutic or *Verstehen* school, as well as functional and teleological modes of explanation.

Like any sweeping and profound social changes, post-communist transformations affect all areas of society. They are "likely to be biological, psychological, demographic, economic, political and cultural—either simultaneously or in succession. Hence, the mechanism of every major social change is likely to be a combination of mechanisms of various kinds coupled together" (Bunge 1997, 417). The systemic framework serves to get these different and interrelated transformation processes into view. The mechanismic mode of explanation helps us to search out and reconstruct the major transformation processes. To be sure, we are not looking for anything like a universal "post-communist transformation mechanism." On the contrary, "concrete, lawful, and scrutable mechanisms are specific, or, if preferred, substance dependent. Hence, there can be no universal explanations of the mechanismic kind. . . . Different kinds of systems, with different mechanisms and under different forces, call for different



explanations" (Bunge 1997, 439-40). "Although all mechanisms are specific (or substrate dependent), it is possible and desirable to group them into large classes on the strength of their similarities" (Bunge 1997, 450). Bunge offers the following definition of a social mechanism:

We define a *social mechanism* as a mechanism in a social system. Since every mechanism is a process in some system, a social mechanism is a process involving at least two agents engaged in forming, maintaining, transforming, or dismantling a social system. There are many types of social system: think, for example, of childless couples and extended families, streetcorner gangs and informal social networks, schools and churches, factories and supermarkets, economies and polities, and local governments and multinational blocs. Correspondingly, there is a large variety of social mechanisms. (Bunge 1997, 447)

The social mechanisms at the center of post-communist transformation are the *major change mechanisms* in the transformation of individual countries. However, we also need to pay special attention to transnational (regional, global) and subnational systems. Thus, for example, the EU as a transnational system in its interaction with transforming countries (national systems), or the Russian regions as subnational systems interacting with the center, are likely to be important elements of major change mechanisms. The concept of change mechanism should not be confused with the concept of system.

Note that our definition presupposes a distinction between system and mechanism: the latter is a process in a system. This distinction is familiar in natural science, where one is not expected to mistake, say, the cardiovascular system for the circulation of the blood or the brain with mental processes. But it is unusual in social studies. (Bunge 1997, 447)

"Mechanism is to system as motion is to body, combination (or dissociation) to chemical compound, and thinking to brain" (Bunge 1997, 449).

One potential source of confusion is the fact that we are dealing with systemic change, so that the major social mechanisms in the transformation occur in systems that are themselves changing—evolving, adapting, or collapsing. Describing fundamental systemic changes in terms of the rate of privatization or liberalization, for instance, is not the same as explanation in terms of major change mechanism. Rate or degree of privatization is an—often overrated



and misinterpreted—indicator of systemic change. By contrast, a major change mechanism involves the concrete system within which this process is taking place (i.e., its particular composition, structure, and environment) and the actual outcome (new, hybrid, or collapsed system).

Thus, marketization refers to a set of major change mechanisms, market reforms are the designed elements in change mechanisms, and the economy is the system (more accurately, one of the systems) that is being changed and changing. In contrast to the voluntaristic and individualistic assumptions that figure prominently in much of the neoliberal transformation literature, the approach presented here does not equate change mechanisms with reform policies. In contrast to the structuralist assumptions that are assigned dominance in much of the critical and culturalist literature, this approach does not reduce change mechanisms to “deeper structural factors.” In the systemic view,

agency is both constrained and motivated by structure, and in turn the latter is maintained or altered by individual action. In other words, *social mechanisms reside neither in persons nor in their environment—they are part of the processes that unfold in or among social systems.* (Bunge 1997, 448; emphasis added)

This conception of social mechanisms also has significant implications for the fundamental problem of the role of social technology in systemic change.

#### Between Social Science and Social Technology: Catalytic Designs

In much of the neoliberal market transition literature, a central assumption is that “the new system” can in fact be designed, since the major mechanisms that maintain a market economy are presumably known and so are the mechanisms (i.e., reforms policies) to put them in place. The two—maintenance and creation of a market economy—are of course not the same. There may be somewhat more knowledge about maintaining established and relatively successful market economies in the West than about creating them under a variety of conditions in the rest of the world. Yet skepticism is called for with respect to the claim that both types of major mechanisms are fully or even sufficiently known. There is the ideological claim that a market system spontaneously creates socially desirable outcomes. In

other words, the social mechanisms that make this particular type of economic system successful are said to be spontaneous as opposed to controlled, economic rather than political, decentralized rather than centralized, as well as intrinsically just (Hayek 1989). This ideological claim, based on empirically untested or untenable assumptions, entails the demand to design and establish an economic system in a controlled, political and centralized fashion that will henceforth maintain itself spontaneously, without the need for political intervention, and in a decentralized way. Describing these ideas in terms of social mechanisms brings out the paradoxical role that design plays in the neoliberal transition literature. Faith in the possibility of rationally designing and establishing a specific type of economic system, on one hand, goes hand in hand with the deep conviction that once established, rational design should not be employed to alter the way the system works, on the other.

I take my own position between the two poles of the neoliberal paradox. I am at the same time more and less sanguine about the role of design in economic and social life. There is both spontaneity and design in most processes of social change and continuity. "Whereas some social systems and their corresponding mechanisms emerge more or less spontaneously, others are designed" (Bunge 1997, 452). Where designs do play an important role, the particular way in which they contribute to social change may contain different mixes of intended, unintended, and perverse effects (Hirschman 1992; Merton 1968, chap. III). To advance our understanding of the actual role of design in post-communist transformation, I conceptualize designs as elements in larger change mechanisms. To underscore this broader conception of design, I speak of *catalytic design* (using the term *catalytic* in its broad sense of an agent that stimulates or precipitates a reaction, development, or change).

Catalytic design is not simply a policy blueprint or institutional template. It may contain that, but it is more (and that "more" is what determines how policy blueprints or institutional transfers work): it is the vision or ideological model; the shifts in political alliances—domestic and international—it builds on or opens up; the potential for redefining state-society relations; the major ways in which society is and is not restructured in the process. It is also less, for even the attempt to faithfully implement a blueprint of a new order can never succeed fully (rational policy fallacy).

The catalytic design is an *emergent property*. As far as catalytic designs for systemic change are concerned, the state (more precisely,

various state actors and institutions) is its major—though clearly not exclusive nor necessarily decisive—agent. This is why we need to pay particular attention to the type and role of the state in its global context. For example, for the Central-Eastern European states, EU integration is the centerpiece of their respective catalytic designs for post-communist political-economic transformation. If we look for a “rational design” at the core of systemic change, it encompasses the broader historical and systemic context plus reform policies and politics (really, the rational design is our model of what is going on, i.e., our rational reconstruction). In this sense, we describe and explain the catalytic design (i.e., model of change) as an emergent property, although clearly not as much of an historical accident as the original emergence of market society and with a greatly more prominent role for human designers and designs.

Finally, I propose to look at the systemic context (domestic and international structures) plus the design (models, templates, and policies) very much as catalysts rather than as a full-fledged plan for systemic change. Catalysts in the sense that they provide only a set of conditions and forces within which processes of social and economic change occur “spontaneously” (in the sense of unplanned and to some extent unintended)—thus, catalytic design. Thus, the same reform measure, change policy, institutional model, or social technology may have different effects in different contexts (e.g., stabilization, privatization). Viewed in conventional terms of design, the policy is assumed to work unless there is something “wrong” in the political, economic, or cultural context in which it is “implemented.” From our alternative viewpoint, a rational design or set of reform policies should not be understood as a simple means to an end that works or doesn’t work. Rather, reform policies, like the general ideologies with which they are associated (e.g., neoliberalism, nationalism, anti-communism, return-to-Europeism), should instead be seen as potential catalysts for change. Whether they are, and of what kind, will depend on the nature of the “reactants” (i.e., the systemic problem context). This is a major point overlooked by conventional voluntaristic and rationalistic designs. This does not commit us to the radical structuralist position according to which individual decisions, events, policies, or plans do not have a decisive impact on the course of change. But just as reduction to structure is unsatisfactory, so is overemphasis of agency and contingency. The mechanistic approach (see above) offers a response to this problem.

In contrast to a policy package such as the neoliberal reform strategy, the concept of catalytic design prominently includes what are usually implicit assumptions about local context, global environment, actors, and problem situations. In one sense, the claim is simple: by using a systemic view of reality, we create the preconditions for identifying most of the relevant change mechanisms. However, this should not be taken as an indication of a hyper-rationalistic design. At least as important as identifying the possibilities and conditions for rationally designed change is the systematic exploration of its limits and the search for catalytic principles (normative),<sup>9</sup> catalytic mechanisms (empirical),<sup>10</sup> and catalytic practices (applied)<sup>11</sup> of change at different levels and in a variety of domains. Thus, when we speak of catalytic designs with respect to a specific case, we attempt to identify (model) the central change mechanisms that catalyze (trigger, bring about, and guide) systemic transformation. Methodologically, we work with the tentative assumption that such models will have relevance for other similar situations, past, present, and future, although to what extent this will be the case is likely to vary from fairly direct applicability to general insight or heuristic.

#### CONCLUDING COMMENT

The above sketch of a social theoretical foundation for transformation studies has responded primarily to a set of philosophical problems—the relationship between social science and social technology and the challenge of cross-disciplinary work. It has taken as its point of departure some fundamental ideas of Popper's social philosophy, examined its strengths and weaknesses in the context of the problems at hand, and offered an alternative approach based on the philosophy of Bunge. Whether and to what extent this alternative is useful for post-communist transformation studies and related problems will be largely determined by theoretical and empirical work. The proof, as so often, is in the pudding.<sup>12</sup>

#### NOTES

1. Arguably, the entire "Third World development" experience could be considered as earlier instances of attempts at controlled, systemic change—though perhaps with

the important caveat that "modernization" did not necessarily entail adoption of Western political and economic institutions. The same could be said about Japan in the 19th century and Turkey in the early 20th century, as well as the experience of all late developers. As Gerschenkron (1962) has shown with respect to 19th-century European economic history, latecomers developed their own set of institutions specific to their local conditions rather than simply copying those of the advanced countries.

2. As borrowed from Roscoe Pound (1922, 99), as Popper (1945, Vol. 1, 210) himself acknowledges.

3. Nationalism, religious fundamentalism, environmentalism, and feminism are alternative, to some extent competing doctrines of social change, although none of these enjoys nearly the same powerful institutional and political support as neoliberalism.

4. As Mario Bunge (1998, 440) has pointed out,

there is a wedge between social science and social technology, namely ideology. This is unavoidable and not deplorable in itself, because technology is neither value-free nor morally neutral. There would be no problem with a proscience and morally right ideology. The trouble is that most ideologies do not meet these conditions.

5. Thus, there is general decline in the political and symbolic attractiveness of neoliberalism as we move from West to East. It is highest in Central Eastern Europe and the Baltic states, more controversial in Southeastern Europe, highly controversial if not simply passé in Russia and the Commonwealth of Independent States, and an anathema in China as well as in the geographic outlier, Cuba.

6. As Janine Wedel (1998) documents, Harvard economist Jeffrey Sachs, for example, advised not only the Polish reform government in 1990 but also appeared in the Polish media to promote the political project of radical transformation.

7. Primarily Anglo-Saxon (United States), secondarily Western European, but very little from East Asia, arguably the most relevant cases for late integration into the capitalist world market.

8. Many of these ideas have been further developed by the leading German critical rationalist, Hans Albert (e.g., 1999a, 1978, 1976). With the exception of a recent English-language collection (Albert 1999b), most of his work is unfortunately not available in English.

9. Example of normative catalytic principles: Western models are best and should be copied, or indigenous models are best.

10. Catalytic mechanisms: the central change mechanisms in a specific case.

11. Catalytic practices: rely on Western experts wherever possible; do without Western experts wherever possible; rhetorically and symbolically talk neoliberal, even if your goals are fundamentally different.

12. The author is currently coordinating a cross-regional comparative study of post-communist transformation, which is following the research program and framework outlined in this article. Support for this project from the Canadian Social Science and Humanities Research Council is gratefully acknowledged.

## REFERENCES

- Agassi, Joseph. 1985. *Technology: Philosophical and social aspects*. Dordrecht and Boston: Reidel.
- . 1987. Methodological individualism and institutional individualism. In *Rationality: The critical view*, edited by J. Agassi and Ian C. Jarvie, 119-50. Dordrecht, the Netherlands: Martinus Nijhoff.
- Albert, Hans. 1976. *Aufklärung und Steuerung*. Hamburg, Germany: Hoffmann und Campe.
- . 1978. *Traktat über rationale Praxis*. Tübingen, Germany: J.C.B. Mohr.
- . 1999a. Die Soziologie und das Problem der Einheit der Wissenschaften. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 51:215-31.
- . 1999b. *Between social science, religion and politics: Essays in critical rationalism*. Amsterdam and Atlanta, GA: Editions Rodopi.
- Aslund, Anders. 1995. *How Russia became a market economy*. Washington, D.C.: Brookings Institution.
- Balcerowicz, Leszek. 1995. *Capitalism, socialism, transformation*. Budapest, Hungary: Central European University Press.
- Bartley, W. W. 1984. *The retreat to commitment*. Peru, IL: Open Court.
- Berend, Ivan T. 1999. *Central and Eastern Europe 1945-1993: Detour from the periphery to the periphery*. Cambridge: Cambridge University Press.
- Bönker, Frank. 1994. External determinants of the patterns and outcomes of East European transitions. *Emergo. Journal of Transforming Economies and Societies* 1 (1): 34-54.
- Bunce, Valerie. 1995. Should transitologists be grounded? *Slavic Review* 54 (1): 111-18.
- Bunge, Mario. 1996a. *Finding philosophy in social science*. New Haven, CT: Yale University Press.
- . 1996b. The seven pillars of Popper's social philosophy. *Philosophy of the Social Sciences* 26:528-57.
- . 1997. Mechanism and explanation. *Philosophy of the Social Sciences* 27:410-65.
- . 1998. *Social science under debate*. Toronto, Ontario: University of Toronto Press.
- Chilton, Patricia. 1995. Mechanics of change: Social movements, transnational coalitions, and the transformation processes in Eastern Europe. In *Bringing transnational relations back in: Non-state actors, domestic structures and international institutions*, edited by Thomas Risse Kappen. Cambridge: Cambridge University Press.
- Dahrendorf, Ralf. 1990. *Reflections on the revolution in Europe*. New York: Random House.
- Dryzek, John. 1990. *Discursive democracy: Politics, policy and political science*. Cambridge: Cambridge University Press.
- Eidlin, Fred. 1997. Blindspot of a liberal: Popper and the problem of community. *Philosophy of the Social Sciences* 27:5-23.
- Fischer, Frank. 1993. Citizen participation and the democratization of policy expertise: From theoretical inquiry to practical cases. *Policy Sciences* 26:165-87.
- George, Susan. 1999. A short history of neo-liberalism: Twenty years of elite economics and emerging opportunities for structural change. Conference on Economic Sovereignty in a Globalising World, 24-26 March, Bangkok.

- Gerschenkron, Alexander. 1962. *Economic backwardness in historical perspective*. Cambridge, MA: Belknap Press of Harvard University Press.
- Gowan, Peter. 1995. Neo-liberal theory and practice for Eastern Europe. *New Left Review* 213:3-60.
- Grabher, Gernot, and David Stark, eds. 1997. *Restructuring networks in post-socialism*. New York: Oxford University Press.
- Greskovits, Bela. 1998. *The political economy of protest and patience: East European and Latin American transformations compared*. Budapest, Hungary: CEU Press.
- Hall, Peter A., ed. 1989. *The political power of economic ideas: Keynesianism across nations*. Princeton, NJ: Princeton University Press.
- Hayek, Friedrich A. 1989. *The fatal conceit: The errors of socialism*. London: Routledge.
- Hirschman, Albert. 1992. *Rival views of market society*. Cambridge, MA: Harvard University Press.
- Holmes, Stephen. 1995. Conceptions of democracy in the draft constitutions of post-communist countries. In *Markets, states, and democracy: The political economy of post-communist transformation*, edited by B. Crawford, 82-100. Boulder, CO: Westview.
- Isaac, Jeffrey C. 1996. The meanings of 1989. *Social Research* 63:291-344.
- Jacoby, Wade. 2002. Talking the talk: The cultural and institutional effects of Western models. In *Postcommunist transformation and the social sciences: Cross-disciplinary approaches*, edited by F. Bönker, K. Müller, and A. Pickel. Boulder, CO: Rowman and Littlefield.
- Jowitt, Ken. 1992. The Leninist legacy. In *Eastern Europe in revolution*, edited by I. Banac, 207-24. Ithaca, NY: Cornell University Press.
- Kabele, Jiri, and Ronald Radzai. 1993. The dynamics of social problems and transformations of Czechoslovak society. *Social Research* 60:763-85.
- Karl, Terry Lynn, and Philippe C. Schmitter. 1995. From an iron curtain to a paper curtain. *Slavic Review* 54:965-79.
- Matzner, Egon, and Ian C. Jarvie. 1998. Papers from the Vienna workshop on Popper's Situational Analysis and the Social Sciences, part I [Special Issue]. *Philosophy of the Social Sciences* 28 (3).
- Merton, Robert K. 1968. *Social theory and social structure*. Enlarged ed. New York: Free Press.
- Müller, Klaus. 1995. Vom Post-Kommunismus zur Postmodernität? Zur Erklärung sozialen Wandels in Osteuropa. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 47:37-64.
- Murrell, Peter. 1992. Conservative political philosophy and the strategy of economic transition. *East European Politics and Society* 6 (1): 3-16.
- . 1995. The transition according to Cambridge, Mass. *Journal of Economic Literature* 33:164-69.
- Myrdal, Gunnar. 1954. *The political element in the development of economic theory*. New York: Simon & Schuster.
- Offe, Claus. 1991. Die deutsche Vereinigung als "natürliches" experiment. In *Experiment Vereinigung: Ein sozialer Grossversuch*, edited by B. Giesen and C. Leggewie, 77-86. Berlin: Rotbuch Verlag.
- Pickel, Andreas. 1989. Never ask who should rule: Karl Popper and political theory. *Canadian Journal of Political Science* 22:84-105.
- . 1992. *Radical transitions. The survival and revival of entrepreneurship in the GDR*. Boulder, CO: Westview.



- . 1995. Official ideology? The role of neoliberal economic reform doctrines in postcommunist transformation. *Polish Sociological Review* 4:361-75.
- . 1997. Neoliberalism, gradualism, and some typical ambiguities and confusions in the transformation debate. *New Political Economy* 2:221-35.
- Pickel, Andreas, and Helmut Wiesenthal. 1997. *The grand experiment: Debating shock therapy, transition theory, and the East German experience*. Boulder, CO: Westview.
- Popper, Karl R. 1945. *The open society and its enemies*. 5th ed. 2 vols. Princeton, NJ: Princeton University Press.
- . 1957. *The poverty of historicism*. 2d ed. London: Routledge Kegan Paul.
- . 1963. *Conjectures and refutations*. 2d ed. New York: Harper & Row.
- . 1970. Normal science and its dangers. In *Criticism and the growth of knowledge*, edited by I. Lakatos and A. Musgrave, 51-58. Cambridge: Cambridge University Press.
- . 1994. *The myth of the framework: In defence of science and rationality*. Edited by M. A. Notturmo. London and New York: Routledge.
- Pound, Roscoe. 1922. *Introduction to the philosophy of law*. New Haven, CT: Yale University Press.
- Poznanski, Kazimierz Z., ed. 1995. *The evolutionary transition to capitalism*. Boulder, CO: Westview.
- Rona-Tas, Akos. 1997. *The great surprise of the small transformation*. Ann Arbor: University of Michigan Press.
- Soros, George. 1998. *The crisis of global capitalism: Open society endangered*. New York: Perseus Books.
- Standing, Guy. 1998. The babble of euphemisms: Re-embedding social protection in "transformed" labour markets. Presented at the Frankfurt Institute of Transformation Studies conference, Postcommunist Transformation and the Social Sciences: Cross-Disciplinary Approaches, 30-31 October, Berlin.
- Sullivan, Michael J. 1994. The impact of Western political thought in Chinese political discourse, on transitions from Leninism, 1986-1992. *World Affairs* 157 (2): 79-91.
- Szacki, Jerzy. 1996. *Liberalism after communism*. Budapest, Hungary: Central European University Press.
- Wedel, Janine R. 1998. *Collision or collusion? The strange case of Western aid to Eastern Europe, 1989-1999*. New York: St. Martin's.
- Wiesenthal, Helmut. 1997. The crisis of holistic policy approaches and the project of controlled system transformation. In *The grand experiment: Debating shock therapy, transition theory, and the East German experience*, edited by A. Pickel and H. Wiesenthal, 91-113. Boulder, CO: Westview.
- Zecchini, Salvatore. 1997. Transition approaches in retrospect. In *Lessons from the economic transition: Eastern Europe in the 1990s*, edited by S. Zecchini, 1-34. Dordrecht, the Netherlands: Kluwer.

*Andreas Pickel is an associate professor of political studies at Trent University, Peterborough, Ontario, Canada. He holds degrees from York University, University of Guelph, and the Free University of Berlin. He is the translator of Karl Popper's Die beiden Grundprobleme der Erkenntnistheorie (The Two Fundamental Problems of the Theory of Knowledge), forthcoming from Open Court Publishers.*



# Explanatory Unification

## Double and Doubtful

USKALI MÄKI

*Erasmus University of Rotterdam, the Netherlands*

Explanatory unification—the urge to “explain much by little”—serves as an ideal of theorizing not only in natural sciences but also in the social sciences, most notably in economics. The ideal is occasionally challenged by appealing to the complexity and diversity of social systems and processes in space and time. This article proposes to accommodate such doubts by making a distinction between two kinds of unification and suggesting that while such doubts may be justified in regard to mere derivational unification (which serves as a formal constraint on theories), it is less justified in the case of ontological unification (which is a result of factual discovery of the actual degree of underlying unity in the world).

### THE ARGUMENT

Explaining much by little, reducing the number of apparently independent phenomena, reducing the number of logically independent lawlike sentences, using the same patterns of argument over and over again to meet different explanatory challenges, economy of thought—these are examples of phrases that are used to express the idea that explanatory unification is a virtue to be pursued in scientific theorizing. Even if it appears as one of the most widely adhered ideals of scientific work, it has not remained entirely unchallenged. By suggesting a distinction between two kinds of unification, I attempt to accommodate one type of doubt about its universal advisability—hence “double and doubtful.”

Three claims will be put forth in this article. First, explanatory unification is widely recognized as a major ideal to be pursued in science. There is no novelty in this claim, yet some documentation will be given to remind the reader that this is indeed so, and that this is so, in particular, also in the science of economics as suggested in an earlier

---

*Received* 21 December 1999

*Philosophy of the Social Sciences*, Vol. 31 No. 4, December 2001 488-506

© 2001 Sage Publications

488

study (Mäki 1990). Yet, this suggestion about unification in economics has been questioned with the descriptive claim that rather than being one of the few disciplines seeking unification, “economics belongs to a family of scientific disciplines” not primarily pursuing explanatory unification (Boylan and O’Gorman 1995, 177). To this is added a claim with a more normative bent: “In view of the fact that unification is not a central preoccupation in many mature sciences, there is no onus on economics to pursue, in the name of mature science, a rigorous policy of unification” (Boylan and O’Gorman 1995, 177). The issue has obvious affinities with recent debates around rational choice in political science, too.<sup>1</sup>

Second, explanatory unification is not a uniform concept but appears in several variants. Two major variants will be identified, to be called “ontological” and “derivational” unification, while further classifications will be put aside. This distinction is not explicit in the well-known discussions of the topic, for example in those by Michael Friedman (1974) and Philip Kitcher (1981, 1989). When I introduced the notion of explanatory unification in the context of economics (Mäki 1990), I focused on the ontological version. I also made the remark that in addition to ontological unification, there is what I then called logical unification and that it is in this latter guise that the idea appears in much of conventional economics (Mäki 1990, 331). In what follows, these two versions will be illustrated with examples from economics and game theory.

Third, a certain worry or skepticism about unification in economics will be accommodated in terms of the above distinction. The suggestion is that there is more justification for the worry if explanatory unification is understood in its derivational version than if it is held in its ontological version. The point is that if there are limits to unification, they had better be ontological in character. One may hope to be able to celebrate unification as a factual discovery, while a more cautious attitude will be recommendable if it is imposed as merely a formal constraint. This argument should have wider applicability—to the rational choice controversies in political science, for example—but this is not pursued in the present article.

#### EXPLANATORY UNIFICATION AS AN IDEAL OF SCIENCE

Explanatory unification is generally acknowledged to constitute a major goal and achievement of the scientific endeavor. It is easy to

document the claim that unification is a very popular idea indeed among practicing scientists at large (for other examples, see Thagard 1978; Kitcher 1989). Antoine Lavoisier (1862), who developed the oxygen theory of combustion replacing the phlogiston theory, appealed to the unifying power of his theory:

I have deduced all the explanations from a simple principle, that pure or vital air is composed of a principle particular to it, which forms its base, and which I have named the oxygen principle, combined with the matter of fire and heat. Once this principle was admitted, the main difficulties of chemistry appeared to dissipate and vanish, and all the phenomena were explained with an astonishing simplicity. (cited in Thagard 1978, 77-78)

In a similar vein, Augustin Fresnel (1866) criticized the particle theory of light and defended the wave theory by claiming that the latter explains a larger range of phenomena:

Thus reflection, refraction, all the cases of diffraction, coloured rings in oblique incidences as in perpendicular incidences, the remarkable agreement between the thicknesses of air and of water which produce the same rings; all these phenomena, which require so many particular hypotheses in Newton's system, are reunited and explained by the theory of vibrations and influences of rays on each other. (cited in Thagard 1978, 78)

We could go on with references to Newton and Darwin, through Maxwell and Bohr, up to the present-day physics and biology. Economics is no exception to the general excitement about explanatory unification. While it is obvious that not all of economics is driven by this ideal, I want to put forth three interrelated claims that I find uncontroversially true: first, much of the most respected parts of economics is motivated by the ideal of unification; second, many developments in economics are celebrated because they are regarded as advancing explanatory unification; and third, the claim that a given theory is not unified and that it does not unify is recognized by large portions of the economics profession as one of the most powerful arguments that can be used against a theory.

The most popular theoretical principles allegedly possessing a lot of unifying power in economics are the familiar two: at the social level, market coordination or the mechanism of demand and supply with equilibrium solutions; and at the level of individual behavior, principles variously put as rational choice, the pursuit of self-interest,

the calculus of pleasure and pain, and maximizing or optimizing under constraints. Economists generally believe that the degree of unification grows as new and diverse phenomena are subsumed under these principles. It is easy to find examples; more will be given in subsequent sections. Here is a beautiful example citing the supply and demand mechanism:

The order, which I have sought to reveal, pervading and moving the most diverse phenomena of the economic world, would be a far less noteworthy and impressive thing were it merely the peculiar product of capitalism. Merchant adventures, companies, and trusts; Guilds, Governments and Soviets may come and go. But under them all, and, if need be, in spite of them all, the profound adjustments of supply and demand will work themselves out and work themselves out again for so long as the lot of man is darkened by the curse of Adam. (Henderson 1924, 17)

Paul Samuelson's seminal *Foundations of Economic Analysis* is explicitly based on the recognition and pursuit of unification. Maximization by economic agents is the key to unification: as this Nobel laureate puts it, "The study of maximizing behavior affords a unified approach to wide areas of current and historical economic thought" (Samuelson [1947] 1983, 23). Underlying this specific conviction there is a general principle:

Most economic treatises are concerned with either the description of some part of the world of reality or with the elaboration of particular elements abstracted from reality. Implicit in such analyses there are certain recognizable formal uniformities, which are indeed characteristic of all scientific method. It is proposed here to investigate these common features in the hope of demonstrating how it is possible to deduce general principles which can serve to unify large sectors of present day economic theory. (Samuelson [1947] 1983, 7)

What is sometimes called the rational expectations revolution in economics more recently is fully in line with Samuelson's program at an abstract level. Accounts of this "revolution" appear to be regularly accompanied by a metatheoretical commentary that justifies and celebrates the theoretical steps taken by appealing to the principle of unification and at the same time denounces contenders for failing to unify. Here is a characteristic statement resorting to the principle of maximizing: "The rational expectations thesis is superior to any competing statement of expectations formation. . . . Alternative theo-

ries . . . fail the one acid test—they do not conform with the basic principles of maximizing behaviour” (Shaw 1984, 105). Robert Lucas, one of the leaders of the movement and winner of the Nobel Prize, puts his achievements in a larger perspective, implying the decisive importance of the principle of unification (based on the doctrine of rational choice microfoundations) as against theoretical fragmentation (attributed by him to Keynesian macroeconomics):

At any given time there will be phenomena that are well-understood from the point of view of the economic theory we have, and other phenomena that are not. We will be tempted, I am sure, to relieve the discomfort induced by discrepancies between theory and facts by saying that ill-understood facts are the province of some other, different kind of economic theory. Keynesian “macroeconomics” was, I think, a surrender (under great duress) to this temptation. It led to the abandonment, for a class of problems of great importance, of the use of the only “engine for the discovery of truth” that we have in economics. Now we are once again putting this engine of Marshall’s to work on the problems of aggregate dynamics. (Lucas 1987, 108)

Many more examples could be given of the importance of the principle of explanatory unification in economics. They would include “local” versions of unification within the traditional domain of economics, comprising a broad range of phenomena from relative prices and industrial organization to unemployment and business cycles. They would also include more “global” versions according to which economists value and pursue the expansion of the scope of economic core theory beyond the traditional boundaries; among the representatives cited would be Gary Becker (family and crime, among other things), James Buchanan and Mancur Olson (politics), and Richard Posner (law and sex).<sup>2</sup> Indeed, anybody familiar with the ways in which economists comment on theories and theoretical advances soon acknowledges the great weight of unification. Given this feature of economics, it is surprising how little has been written explicitly about it in the philosophy and methodology of economics (see Mäki 1990, 1992, 1997a; Boylan and O’Gorman 1995, 171-77; Kincaid 1997, 100-8). On the other hand, given the popularity of the idea of unification across fields of science and their developments, it is not surprising that philosophers of science at large have paid attention to this principle. Some of them have made it the cornerstone of their accounts of what good scientific explanation is like. We find it in William Whewell’s idea of consilience in his *Philosophy of the Inductive Sci-*

*ences*. In the recent years, we find it in the work by philosophers such as Michael Friedman, Paul Thagard, Philip Kitcher, and others.

Here are two remarks of clarification before proceeding to more specific expositions. First, unification can be viewed as being a matter of the explanantia/explananda ratio. This way of putting the idea is flexible enough to permit a variety of further specifications. In particular, it is consistent with both the derivational and ontological versions, depending on the character of the explanantia and explananda: linguistic, propositional, ontic. Irrespective of the specification, what we have here is a kind of input-output ratio that measures what an economist might call “explanatory efficiency.” Second, *unify*, *unified*, and *unification* are terms that are used in connection to both theories and phenomena. We may say that a theory is unified and also that phenomena are unified by a theory. A unified theory is one that includes a small coherent and organically connected set of explanatory principles that is available for systematic and consistent application to a variety of phenomena.<sup>3</sup> On the other hand, phenomena are unified with one another when they are explained in terms of one and the same set of explanatory principles. Thus, so used, *unification* pertains both to the explanantia and to the explananda of a science. Our examples given in the present section were mixed regarding their focus with respect to these two possibilities.

#### DERIVATIONAL UNIFICATION

Combining the two ideas that explanation is a matter of inference or derivation and that explanation involves unification of phenomena gives us the notion of derivational unification. Michael Friedman’s early account of explanatory unification starts with the suggestion that “this is the essence of scientific explanation—science increases our understanding of the world by reducing the total number of independent phenomena that we have to accept as ultimate or given” (Friedman 1974, 15). He then proceeds to express this general idea in sentential terms: explanation is a matter of reducing the number of logically independent lawlike sentences (Friedman 1974, 15-18); as Kitcher later put the idea, explanatory unification is a matter of establishing the best trade-off between minimizing the number of premises and maximizing the number of conclusions of explanatory arguments (Kitcher 1989, 431). This is also roughly the way I charac-

terized what I called logical unification: "Logical unification is brought about when more and more statements within a discipline become derivable from the same set of axioms, or when the same set of statements becomes derivable from a smaller set of axioms" (Mäki 1990, 331). This notion can as well be called derivational unification, since it is "derivational efficiency" that is referred to. Explanations are basically understood as arguments, as derivations from premises to conclusions, where the conclusions give the explananda and the premises give the explanantia.

Kitcher's account suggests to use "patterns of derivation" as the unit instead of pairs of premises and conclusions: understanding the phenomena is a matter of "seeing connections, common patterns, in what initially appeared to be different situations" (Kitcher 1989, 432). Explanatory unification becomes a matter of establishing a trade-off between minimizing the number of derivational patterns and maximizing the number of conclusions. Derivational patterns are abstract schemes, instantiated in specific applications. Such applications require following a set of "filling instructions" that give directions for replacing the dummy letters in the schematic sentences that are included in the schematic patterns. The idea of scientific progress can then be characterized in these terms:

Science advances our understanding of nature by showing us how to derive descriptions of many phenomena, using the same patterns of derivation over and over again, and, in demonstrating this, it teaches us how to reduce the number of types of facts we have to accept as ultimate (or brute). (Kitcher 1989, 432; italics deleted)

This seems to be a more promising way of characterizing derivational unification as it appears in economics. Invoking the same pattern of derivation over and over again—this is a prominent idea that an economist easily recognizes in his own discipline. Constrained maximization is the case in point. Paul Samuelson's *Foundations of Economic Analysis* is a prime example of work that is explicitly based on this idea. The book begins with these words:

The existence of analogies between central features of various theories implies the existence of a general theory which underlies the particular theories and *unifies them* with respect to those central features. . . . It is the purpose of the pages that follow to work out [the implications of this fundamental principle] for theoretical and applied economics . . .

seemingly diverse fields—production economics, consumer’s behavior, international trade, public finance, business cycles, income analysis—possess striking formal similarities. . . . Only after laborious work in each of these fields did the realization dawn upon me that *essentially the same inequalities and theorems appeared again and again*. (Samuelson [1947] 1983, 3; italics added)

Indeed, it is no news to a student of economics to be pointed out the fact that much of economic theorizing is a matter of invoking the same derivational patterns over and over again, irrespective of the special field of study. Samuelson’s formulation is in terms of maximization and equilibrium, those paradigmatic elements of what is known as the conventional economic approach to the study of society and human behavior:

The general method [which lies at the bottom of much of economic theory] may be very simply stated. In cases where the equilibrium values of our variables can be regarded as the solutions of an extremum (maximum or minimum) problem, it is often possible regardless of the number of variables involved to determine unambiguously the qualitative behavior of our solution values in respect to changes of parameters. (Samuelson [1947] 1983, 21; italics deleted)

Without going into details, it seems to me that Samuelson’s program, shared by much of the mainstream of economics ever since, fits by and large with Kitcher’s description. There is the common derivational pattern, there is implied a set of filling instructions needed to apply the pattern to a diversity of cases so as to establish that “essentially the same inequalities and theorems appeared again and again,” from consumer behavior to business cycles and international trade.

Here is another example. Robert Aumann (1985), in his “What Is Game Theory Trying to Accomplish?” gives a fairly unambiguous statement of what I mean by “derivational unification” without ontological grounding. Aumann states that the basic aim of science is understanding. In the familiar “unificationist” manner, he suggests that scientific understanding has three components. The first is

fitting things together . . . relating, associating, *recognizing patterns*. Snowflakes are hexagonal; the shells of certain snails are logarithmic spirals; buses on busy routes arrive in bunches; in their orbits around the sun, the planets sweep out equal areas in equal times.



The second component is *unification*:

The broader the area that is covered by a theory, the greater is its "validity." . . . Part of the greatness of theories like gravitation or evolution, or the atomic theory of matter, is that *they cover so much ground, that they "explain" so many different things.* (italics added)

The third component is *simplicity*. It can be divided into subcomponents. One of them is sparseness in the number of exogenous parameters. Another is "sparseness in the basic structure of the theory." Yet another is

simplicity in the sense that is opposite to difficulty. For a theory to be useful working with it must be practical. If you cannot figure out what it implies, it won't unify anything, it won't establish relationships. (Aumann 1985, 29-31)

Aumann then proceeds to suggest that the above characterization of the notion of understanding has obvious implications regarding the dispensability of the concept of truth. We can recognize a strong instrumentalist flavor in the following sentences:

Most readers will by now have understood that, in my view, scientific theories are not to be considered "true" or "false." In constructing such a theory, *we are not trying to get at the truth, or even to approximate to it: rather, we are trying to organize our thoughts and observations in a useful manner.* One analogy is to a filing system in an office operation, or to some kind of complex computer program. We do not refer to such a system as being "true" or "untrue"; rather, we talk about whether it "works" or not, or, better yet, how well it works. . . . Similarly, scientific theories must be judged by how well they enable us to organize and understand our observations; by how well they "work." (Aumann 1985, 31-32; italics added)

In the passages cited above, Aumann puts forth general statements about scientific theories, understood instrumentalistically, providing unification of phenomena. He deals with game theory as a special case of this general account by commenting on the solution concepts:

People ask, since game theory offers a multiplicity of solution notions, what good can it be? Which solution notion is the right one? How do people "truly" behave? . . . None of the solution notions tells us how people truly behave. They do not go about organizing blocking coalitions.

tions, as the core might suggest; they do not object and counter-object as in the bargaining set; they do not declare dividends as in Harsanyi's value; and so on. Rather, *a solution notion is the scientists' way of organizing in a single framework many disparate phenomena and many disparate ideas.* (Aumann 1985, 34-35; italics added)

Turning then to a commentary of economics more narrowly, Aumann once again refers to the twin ideas of derivational unification and the irrelevance of truth. He says that

the validity of utility maximization does not depend on its being an accurate description of the behavior of individuals. Rather, it derives from its being the underlying postulate that *pulls together most of economic theory*; it is the major component of a certain way of thinking, with many important and familiar implications, which have been part of economics for decades and even centuries. (Aumann 1985, 35; italics added)

In a manner which most economists immediately recognize as familiar, Aumann goes on by appealing to derivational unification as a constraint that rivals to maximization fail to meet:

Alternatives such as satisficing have proved next to useless in this respect. While attractive as hypotheses, there is little theory built on them; *they pull together almost nothing*; they have few interesting consequences. In judging utility maximization, we must ask not "Is it plausible?" but "*What does it tie together, where does it lead?*" (Aumann 1985, 35; italics added)

It is obvious from the above passages that Aumann's account presents unification as a derivational accomplishment without ontological groundings. In this respect it resembles Kitcher's "Kantian" account. This feature of Kitcher's exposition of the unification view of explanation is manifest in his treatment of causality. Explanation is not a matter of describing causal relations in the world, it is rather the other way around: causal relations are a function of explanatory relations. "What is distinctive about the unification view is that it proposes to ground causal claims in claims about explanatory dependency rather than vice versa" (Kitcher 1989, 436). The notion of causality does not enjoy any autonomously ontological, nonepistemic status:

there is no sense to the notion of causal relevance independent of that of explanatory relevance and . . . there is no sense in the notion of explana-

tory relevance except that of figuring in the systematization of belief in the limit of scientific inquiry, as guided by the search for unification. (Kitcher 1989, 499)

It is noteworthy that Kitcher thinks his Kantian impositionism is a built-in characteristic of the unification view—an idea that will be disputed next.

### ONTOLOGICAL UNIFICATION

Explanatory unification in the derivational mode is an option, but Kitcher's suggestions notwithstanding, we should not be misled to thinking that it is the only option. It is not, as will be documented in this section: practicing researchers do entertain the idea of ontological unification as well. First, we need a rudimentary idea of what ontological unification is. In contrast to derivational unification, ontological unification is based on the referential and representational capabilities of theories, while derivational unification is based on their inferential capabilities (Mäki 1990, 331). Ontological unification is a matter of redescribing apparently independent and diverse phenomena as manifestations (outcomes, phases, forms, aspects) of one and the same small number of entities, powers, and processes. Those phenomena are thereby revealed to be only apparently independent; as a matter of actual fact, they are dependent on the same underlying structure of entities, forces, and processes (Aronson 1984). The notion of ontological unification, unlike that of mere derivational unification, is presumed to include some deeper idea of why exactly unification would be a virtue worth pursuing.

If there is unity among a set of phenomena, it is a matter of their sharing the same ontic foundations (causes, origins, constituents). Unity among phenomena is a matter of what they are and how they come about, and it is a matter of discovery rather than imposition to establish this. Discovery is a matter of inventing a theory that correctly picks out the relevant entities and represents their relevant properties. A paradigm case is the unifying power of Newtonian mechanics. A huge variety of sublunary and superlunary phenomena can be represented as having the same causes: planets and falling apples, tides and trajectories of cannon balls, galactic constellations and molecular formations—they all manifest the same forces of gravitation. The degree of ontological unification is not a sentential ratio or

a pattern/conclusion ratio as in the case of derivational unification but rather the ratio between the number of kinds of entities (forces, fields, processes) referred to by a theory and the number of kinds of phenomena successfully redescribed as manifestations (phases, aspects, forms, effects) thereof.

The above presumptions suggest some of the relevant contrasts between the two kinds of unification: ontological as opposed to mere derivational unification gives priority to entities rather than sentences; reference and representation rather than inference and derivation; discovery rather than imposition. Note that ontological unification and mere derivational unification are supposed to be so contrasted—but putting the suggestion in this way leaves room for the possibility that derivational and ontological unification coincide, or perhaps that derivational unification has partial ontological grounds. My hunch is that this is a contingent issue; there is no necessity for the two kinds of unification to be related in one particular way or another.

I claimed that ontological unification is an option entertained by practicing scientists and promised to provide documentation. My examples are taken from economics. No more than very brief and suggestive illustrations will be given (a more detailed case study is given in Mäki 1990). My first example is Milton Friedman and his famous 1953 methodological essay. This will surprise some readers, since Friedman is usually identified as a nonrealist, and the notion of ontological unification is a realist idea. As argued elsewhere, Friedman is not consistent, and major parts of his essay can, contrary to conventional readings, be construed as a realist defense of his favorite economics (Mäki 1992). Here is one key passage:

A fundamental hypothesis of science is that appearances are deceptive and that there is a way of looking at or interpreting or organizing the evidence that will reveal superficially disconnected and diverse phenomena to be manifestations of a more fundamental and relatively simple structure. (Friedman 1953, 33)

The key ideas characterizing ontological unification are present in this passage in rough outline. There is the idea of there being “superficially disconnected and diverse phenomena” to be unified. There is the idea of “a more fundamental and relatively simple structure” in terms of which unification can be accomplished. And there is the idea that unification is a matter of showing that those disconnected phe-

nomena are only apparently disconnected, because the facts of the matter are such that these phenomena are “manifestations” of one and the same “fundamental and relatively simple structure.” Unification, according to this picture, is not just a matter of derivational success but rather a matter of successfully representing how things are related in the causal order of things in the world.

It is a different issue to ask what contents should be ascribed to the “fundamental and relatively simple structure” in Friedman’s scheme for it to help unify apparently diverse phenomena. Two suggestions have been made in the literature. First, in the article where I suggested this reading of Friedman (as subscribing to ontological unification as a goal of economics), I also suggested as one possibility that depicting economic agents as striving for maximum expected returns might be believed by Friedman to capture such a structure capable of unification (Mäki 1992). He might believe that a broad range of diverse phenomena can be accounted for in terms of a single mental fact (plus, of course, a number of auxiliary factors relevant to those diverse cases, respectively). Second, Jack Vromen (1995, 198-200) has suggested a second reading on which the mental states and processes of economic agents are substituted for impersonal market forces as constituting “the fundamental structure.” On this reading, maximization of expected returns is just a shorthand for the conditions of survival in economic “natural selection.” Whatever the correct reading—supposing there is such a thing as a correct reading in this case—of Friedman is on this specific detail, the piece of textual evidence invoked above may be taken to show that, as a general metatheoretical principle, he subscribes to the ideal of ontological unification. This does not easily fit with the received image of Friedman’s methodological views.

Consider another example, the case of Austrian economics. It permits a detailed analysis of ontological unification, and it does so perhaps more easily than Friedman’s case (Mäki 1990). Subjective valuation and purposive individual action as well as their systematic unintended consequences—these are the fundamental realities by reference to which explanations are designed. Economic phenomena and institutions are explained by theoretically redescribing them as forms or manifestations of such realities. Such theoretical redescriptions can be construed in terms of ontological identification statements analogous to the familiar ones of physics and chemistry, such as “Water is  $H_2O$ ” and “Temperature of a gas is the mean kinetic energy of individual gas molecules” (Mäki 1990, 324):

- [A] Social entities are aggregates or averages of individual entities, these latter entities being invested with meaning by acting individuals.
- [C] Social entities are unintended consequences of actions by human individuals.

These two statements give two principles of the constitution of social entities, the aggregative and the causal. The meaning of *are* in [A] and [C] is not supposed to designate mere coextensionality of the two descriptions. What *are* is intended to mean here can be expressed by phrases such as “are really,” “are nothing but,” “are ultimately,” and so forth. This makes the relationship between the two sides of *are* asymmetrical. Carl Menger’s theory of imputation, his account of the value of the inputs of production—goods of higher order—involves the idea of ontological identification: “The value of goods of higher order is, therefore, in the final analysis, nothing but a special form of the importance we attribute to our lives and well-being” (Menger [1871] 1976, 152; for a detailed account of Menger’s underlying philosophy, see Mäki 1997b).

It is notable that identification statements such as [A] and [C] provide ontological variants of explanatory patterns that are being applied over and over again to a vast array of phenomena (Mäki 1990, 326). Such patterns differ from the ontologically ungrounded derivational patterns in Kitcher’s account in that they involve the conviction or conjecture that [A] and [C] and other such identification statements are true of the way the world is constituted and that such convictions or conjectures guide the search for explanations. Yet, they alone do not imply any precise derivational patterns.

Ontological unification then is a matter of invoking ontological identifications such as [A] and [C] or their specific forms in a number of cases. Menger subscribed to this idea when stating “that all phenomena of value are *the same in nature and origin*, and that the magnitude of value is always governed according to the same principles” (Menger [1871] 1976, 173; italics added). Many later Austrians tend to share this thrust, as witnessed by Kirzner’s statement: “Exactly *the same* competitive-entrepreneurial market *process is at work* whether it manifests itself through prices adjusting toward general (or partial) equilibrium patterns or through the adjustment of commodity opportunities made available, techniques of production, or the organization of industry” (Kirzner 1973, 129; italics added). In these passages, the italicized expressions highlight the ontological character of the suggested unifications.

Explanatory progress can then be defined in terms of increasing degree of ontological unification. Here is Hayek's famous passage: "It is probably no exaggeration to say that every important advance in economic theory during the last hundred years was a further step in the consistent application of subjectivism" (Hayek 1955, 31). This can be rephrased as the claim that an increasing range of phenomena has been discovered to be forms or manifestations of subjective valuations and actions by individual economic agents. Ontological unification is a matter of factual discovery.

#### A WORRY ABOUT UNIFICATION

Unification is conventionally listed as one of the virtues of scientific theories. Yet, there are worried voices raising questions about whether explanatory unification is a virtue after all. Here is an example putting forth an argument against unification in economics:

Since economic time is historical, rather than logico-mathematical, there is a grave danger that a single-minded quest for unification, through the use of mathematical models, could conceal from view the rich diversity of economic conditions which contributed to the birth, development or demise of sundry contingent economic institutions and structures dispersed across the history of divergent human cultures and civilizations. (Boylan and O'Gorman 1995, 177)

Now it is easy to agree on the observation concerning "the rich diversity of economic conditions" and "divergent human cultures" and the like, but this cannot serve as an argument against unification. There is no justified argument from the observation of diversity to the denial of unity and of the pursuit of unification. The point of explanatory unification is exactly to redescribe such diversities and divergencies as something else, as manifestations of underlying unities. Explanatory unification is a matter of turning apparent diversities into real unities as it were. Diversity and unification can go together.

Let me try to translate the worry into what I believe is a more justified statement:

There are numerous diversities in the economies across the history of human cultures. Whether there are in addition some unities underlying such diversities is a matter of contingent fact, to be discovered by way



of research. A single-minded quest for a single kind of unification, namely derivational unification, should be discouraged. An open-minded quest for ontological unification, that is, discovery of underlying unities, should be encouraged. There may be limits to unification, but they are based on whatever unity there is in the world.

My view is that if one feels like objecting to the ideal of explanatory unification in a discipline like economics, one had better try to show that there is some major shortcoming in economics (or current economics) because of which ontological unification is not forthcoming. This would require two moves: first, the adoption of the ontological variant of unification; second, an argument to the effect that the prerequisite of ontological unification, namely the successful identification of the widely recurring major elements in the way the world works, is not met.

One has to be clear about what will have been accomplished through such an argument. One will have shown that even if derivational unification may have been achieved, ontological unification remains unattained. One may then suggest that this is no special recommendation for the theory. One may even seek to argue that sticking to a theory that is only able to exhibit derivational unification serves as an obstacle to further progress in economics. There is both a descriptive and a normative aspect in the argument thus far. Consistent with this, nothing would prevent us from adopting the notion of ontological unification as a normative ideal or theoretical virtue that should guide and constrain economic theorizing.

To clarify the point further, we may distinguish between “unification as formal constraint” and “unification as factual discovery.” Consider the example of self-seeking maximization in economic theory. When functioning as a constraint, the principle of unification prescribes that the results of theoretical work be derivable from the assumption of maximization by economic agents—deviations from this rule will be proscribed as *ad hoc*. On the other hand, when viewed as a discovery, unification would be a matter of finding out that the way the world works is such that self-seeking maximization indeed underlies a large variety of types of phenomena. We may further suggest that it is in its derivational guise that unification easily adopts the role of a formal constraint and that in its ontological guise its primary role may be that of factual discovery. (This should not be taken to exclude the obvious possibility that unification may serve both purposes at the same time.) One may then justifiably object to using the principle of unification only as a formal constraint, based on the

derivational properties of certain theoretical assumptions. This objection should be particularly well taken in a discipline such as economics in which the empirical control of theorizing is rather weak. It would be motivated by the reasonable worry that using derivational unification as a logical constraint on theorizing would further immunize it from empirical criticism.

The notion of ontological unification has one advantage over the notion of derivational unification. The advantage is this: the power of a theory to unify may be thought to have limits that are based on the degree of ontic unity of its domain.<sup>4</sup> In this picture, factual inquiry into this domain and its boundaries adopts a special role. As the conclusion of such factual inquiry, we may draw such boundaries variously. Naturally, scientists are inclined to generalize on past discoveries and take those generalizations as guidelines for future inquiries. But overall, such guidelines, and the ideal of explanatory unification itself, should have an a posteriori character. Another way of putting this thought is to say that unification is contingent upon factual discoveries about causal structures in the world rather than being constitutive of explanation.

This leads to one final point. It is indeed sometimes suggested or implied that unification is a constituent or defining characteristic or necessary condition of scientific explanation, or at least of good scientific explanation: to explain is to unify. This is not how I see it. I would rather think of the power to (ontologically) unify as a virtue of theories that are explanatory due to other accomplishments—mainly thanks to tracing the causal trajectories of phenomena. Unifying power may thus serve a purpose as a criterion of justification. Increased understanding may not as such amount to increased unification, but increased unification may provide grounds for increased reliance on the capacity of a theory to provide understanding. Explanatory understanding is a matter of revealing the way the world works, without presupposing a priori that the world itself is maximally unified or unified to any particular degree.

#### NOTES

1. Green and Shapiro (1994) identify explanatory unification (they call it universality) as a dominant pursuit of rational choice theorizing in political science. They are bothered by it but do not want to go as far as proscribing it in general. Their recipe is

rather to regulate and constrain the pursuit of universality. My argument in the present article shares this general thrust. See also Mäki (2000).

2. On the rationale of, and constraints on, these more radical versions (or what is known as economics imperialism), see Mäki (2000). We might add to the above list of “expansionist” economists such philosophers as Nicholas Rescher and Alvin Goldman who have offered accounts of knowledge and science in economic terms.

3. The notion of unified theory has turned out to be difficult for philosophers to elaborate in detail. According to one proposal, a unified theory satisfies the “organic fertility requirement” that requires that the whole theory has more testable content than the sum of the testable contents of its parts (Watkins 1984, 205).

4. For an important account that may be used for developing the principle that explanatory unification be constrained by the actual degree of ontic unity in the world, see Dupré (1993).

## REFERENCES

- Aronson, Jerrold. 1984. *A realist philosophy of science*. London: Macmillan.
- Aumann, Robert J. 1985. What is game theory trying to accomplish? In *Frontiers of economics*, edited by K. Arrow and S. Honkapohja. Oxford: Blackwell.
- Boylan, Thomas, and Paschal O’Gorman. 1995. *Beyond rhetoric and realism in economics*. London: Routledge.
- Dupré, John. 1993. *The disorder of things. Metaphysical foundations of the disunity of science*. Cambridge, MA: Harvard University Press.
- Friedman, Michael. 1974. Explanation and scientific understanding. *Journal of Philosophy* 71:5-19.
- Friedman, Milton. 1953. *Essays in positive economics*. Chicago: University of Chicago Press.
- Green, Donald, and Ian Shapiro. 1994. *Pathologies of rational choice theory*. New Haven, CT: Yale University Press.
- Hayek, Friedrich. 1955. *The counter-revolution of science*. Glencoe: Free Press.
- Henderson, Hubert D. 1924. *Supply and demand*. London: Nisbet.
- Kincaid, Harold. 1997. *Individualism and the unity of science*. Lanham: Rowman & Littlefield.
- Kirzner, Israel. 1973. *Competition and entrepreneurship*. Chicago: University of Chicago Press.
- Kitcher, Philip. 1981. Explanatory unification. *Philosophy of Science* 48:507-31.
- . 1989. Explanatory unification and the causal structure of the world. In *Scientific explanation*, edited by P. Kitcher and W. Salmon, 410-506. Minneapolis: University of Minnesota Press. (*Minnesota Studies in the Philosophy of Science*, Vol. 13)
- Lucas, Robert. 1987. *Models of business cycles*. Oxford: Blackwell.
- Mäki, Uskali. 1990. Scientific realism and Austrian explanation. *Review of Political Economy* 2:310-44.
- . 1992. Friedman and realism. *Research in the History of Economic Thought and Methodology* 10:171-95.
- . 1997a. The one world and the many theories. In *Pluralism in economics*, edited by Andrea Salanti and Ernesto Screpanti, 37-47. Cheltenham: Edward Elgar.

- . 1997b. Universals and the *Methodenstreit*: A reexamination of Carl Menger's conception of economics as an exact science. *Studies in History and Philosophy of Science* 28:475-95.
- . 2000. Economics imperialism: Concept and constraints. Unpublished manuscript.
- Menger, Carl. [1871] 1976. *Principles of economics*. New York: New York University Press.
- Samuelson, Paul. [1947] 1983. *Foundations of economic analysis*. Enlarged ed. Cambridge, MA: Harvard University Press.
- Shaw, G. K. 1984. *Rational expectations*. Brighton: Wheatsheaf.
- Thagard, Paul. 1978. The best explanation: Criteria for theory choice. *Journal of Philosophy* 75:76-92.
- Vromen, Jack. 1995. *Economic evolution*. London: Routledge.
- Watkins, John. 1984. *Science and scepticism*. Princeton, NJ: Princeton University Press.

*Uskali Mäki is a professor of philosophy at the Erasmus University of Rotterdam, the Netherlands. He is an editor of the Journal of Economic Methodology and the Handbook of Economic Methodology. His recent publications include "Mechanisms, Models, and Free Riders," in Economics and Methodology: Crossing the Boundaries, edited by R. Backhouse, D. Hausman, U. Mäki, and A. Salanti (Macmillan, 1998); "Science as a Free Market: A Reflexivity Test in an Economics of Economics," in Perspectives on Science (April 1999); "Reclaiming Relevant Realism," in Journal of Economic Methodology (March 2000); and The Economic World View: Studies in the Ontology of Economics, ed. (CUP 2001).*

# “Agency” as a Red Herring in Social Theory

STEVEN LOYAL

*University College Dublin*

BARRY BARNES

*University of Exeter, England*

The central argument of this article is that there is no fact of the matter, no evidence, however tentative or questionable, that will serve adequately to identify actions “chosen” or “determined” for the purposes of sociological theory. This argument will be developed with reference to the two theorists of the greatest importance in advocating the sociological value of the concept of agency: Talcott Parsons, with his “voluntaristic theory of action,” set the scene for the whole agency and structure debate in modern sociology, and Anthony Giddens, in his theory of structuration, provides the most comprehensive recent account. Both theorists put forward grounds and justifications for their use of the concepts of “choice” and “agency,” but it will be argued here that in the last analysis, none of them has any sociological merit.

The concept of agency occupies a central position in much current social theory, wherein it is employed in many and various ways. The present article is concerned only with that part of the literature that specifically refers to human agency. In this literature, the concept of agency is contrasted with that of social structure wherefrom it derives its meaning relationally. “Agency” stands for the freedom of the contingently acting subject over and against the constraints that are thought to derive from enduring social structures. To the extent that human beings have agency, they may act independently of and in opposition to structural constraints, and/or may (re)constitute social structures through their freely chosen actions. To the extent that they lack agency, human beings are conceived of as automata, following the dictates of social structures and exercising no choice in what they do. That, at any rate, is the commonest way of contrasting agency and structure in the context of what has become known as the structure/agency debate.

---

*Received* 29 February 2000

*Philosophy of the Social Sciences*, Vol. 31 No. 4, December 2001 507-524

© 2001 Sage Publications

Whatever opinion is taken of the quality of this recent debate, it is clear that through the issues it raises it resonates with a host of other important questions. The notion of a freely acting individual has had a central place in classical liberal theory since the late 16th and early 17th centuries as a crucial element in the very idea of capitalism. Equally, "free" actors, able to transform their surroundings through active intervention, can be of theoretical value in the context of a socialist politics (Loyal forthcoming). More generally, accounts of human agency can raise issues of individual responsibility that cross and transcend right/left political divisions (Barnes 2000).

A contrast of "freely chosen" and "determined" action has often been put to political use, and it is no part of our objectives in this article to question the ways in which it has been so used. Our present concern is simply with the sociological utility of the concept of agency. And the argument will be that it has none, that there is no fact of the matter, no evidence, however tentative or questionable, that will serve adequately to identify actions as "chosen" or "determined" for the purposes of sociological theory. This argument will be developed with reference to the two theorists who are of the greatest importance in advocating the sociological value of the concept of agency: Talcott Parsons, with his "voluntaristic theory of action," set the scene for the whole agency/structure debate in modern sociology, and Anthony Giddens, in his theory of structuration, provides the most comprehensive recent account. Both theorists put forward grounds and justifications for their use of concepts of "choice" and "agency," but it will be argued here that, in the last analysis, none of them has any sociological merit.

#### PARSONS'S VOLUNTARISM

In *The Structure of Social Action*, Parsons (1937) set out what became a key reference point for all further analyses of action within social theory. Central to the analysis was his account of a "unit act":

1. The act implies an agent, an actor.
2. The act must have an end, a future state of affairs toward which the action is orientated.
3. The act must be initiated in a situation in which intervening action is necessary to bring about the state of affairs that is the actor's end. This situation is in turn analyzable into two kinds of elements: those over

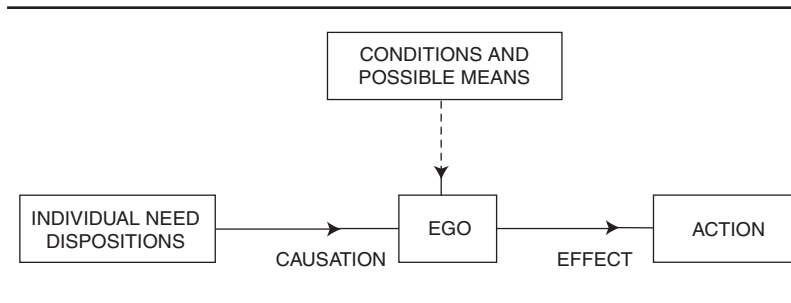


Figure 1

which the actor has no control and those over which she has control. The former are conditions of action and the latter the means of action.

4. The means and ends of action are to be understood by reference both to *individual* factors (wants or need dispositions in the case of ends, individual rational calculations in the case of means) and to a *social, normative* element involved in their constitution.

Let us consider first of all an account that refers to just the first three aspects of the unit act. The actor seeks to realize ends in a situation wherein there are given material conditions to be taken into account and possible physical means of realizing the ends available. The situation is thus understood much as a modern rational choice theorist might understand it, and a rational choice theorist would happily go on to speak of an actor rationally calculating the best thing to do in the situation and acting accordingly. Parsons himself of course was well aware of this way of understanding an act. He called it a "utilitarian" approach, and he sometimes adopted something very like it himself. Human beings, he believed, had given "egoistic" ends (wants/desires/innate need dispositions) and could be prompted into action by them. But Parsons was reluctant to cede much explanatory scope to a utilitarian approach, which he regarded as "reductive." The ends or wants that prompted action here, while internal to the individual body, were nonetheless external to the individual's acting self or ego and operated causally upon it. The result was not true action at all but rather something analogous to animal behavior in being determined and not chosen (see Figure 1).

Parsons was intuitively averse to a reductive utilitarian approach to the explanation of action, but he also had a powerful argument to advance against it: if egoistic ends caused actions, there could be no social order, such as we manifestly observe. Parsons cited Hobbes's

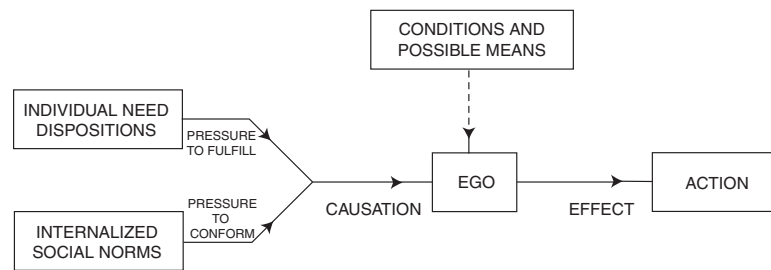


Figure 2

famous argument here, that human beings who act simply to fulfill egoistic desires will merely produce “a war of all against all.” Such a state of war can only be overcome, claimed Parsons, and a peaceful social order established, if the egoism of individuals is overridden. Evidently, that egoism is overridden, but by what? This is where the fourth aspect of action mentioned by Parsons enters: the actor has a normative orientation; she acts in relation to social norms as well as individual desires.

Parsons believed that shared norms and values were internalized into the individual during socialization and came to constitute an alternative basis for action to that offered by individual desires. Insofar as enough action was oriented to social norms rather than individual desires, a social order could be enacted and a Hobbesian war avoided. The individual, caught between the urges of egoism and the prompting of internalized norms (both internal to the individual but external to the acting self or ego) only had to act sufficiently often in relation to the latter for Hobbes’s problem to be solved.

How individual ends and social norms press upon the center of action in the individual could perhaps be represented by a simple extension of our initial figure as in Figure 2, wherein the actor is caused to act in the way required by the stronger of two opposed pressures.

But the idea of the individual’s acting in the direction of the stronger cause is just as reductive as that of the individual’s acting in response to a single cause, and Parsons preferred instead to represent matters as in Figure 3.

Figure 3 summarizes Parsons’s voluntaristic theory of action. At the same time as adding social norms to individual ends as factors impinging on action, Parsons took the opportunity to change the relationship between the factors and the action from one of causation to



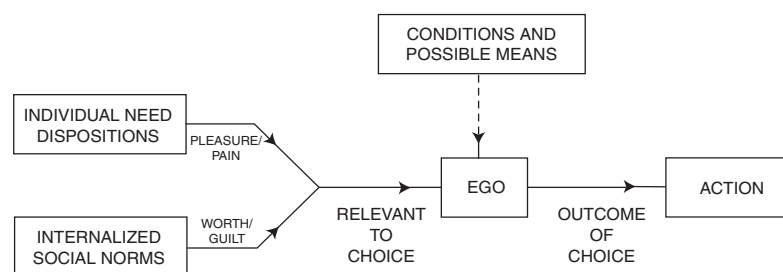


Figure 3

one of choice. The actor cannot choose the pains and pleasures associated with action but can choose how far to take account of them in acting. Naturally, as the pain of deviating from norms increases, the actor will choose conformity more frequently, but it is choice nonetheless that results in action. Action is now voluntary, not determined, with the choosing agent placed between, as it were, two kinds of pressure:

Action must always be thought of as involving a state of tension between two different orders of elements, the normative and the conditional. (Parsons 1937, 732)

Parsons recognized, of course, that individual actors are moved to conform to norms by external as well as internal pressures. The sanctions of others will press the individual to conform to norms. But these sanctions, which the individual experiences as among the external conditions of action, are merely the products of the internal pressure of norms upon others: sanctions are secondary and derivative supports of normative order and have no independent significance. They do no more than intensify and extend the scope of the power of internalized norms: "The principal basis . . . of the efficacy of a system of rules as a whole lies in the moral authority it exercises. Sanctions form only a secondary support" (1937, 402). And again, "the primary source of constraint lies in the moral authority of a system of rules. Sanctions thus become a secondary mode of enforcement of the rule, because the sanctions are, in turn, dependent on moral authority" (1937, 463).

Thus, in the last analysis, for Parsons, the amount of action generated in conformity with a norm will vary systematically according to how strongly and extensively the norm is internalized, how strong is the desire to realize individual ends that oppose it, and how much

work and effort is entailed in conforming to the norm. According to his voluntaristic account (see Figure 3), people will freely choose how to act while taking into account all of these things. But notice here that all of them can also be thought of bearing on action as causes, with what is done being determined as that which maximizes the net cost/benefit of action, as in Figure 2. And above all here, note how there is no evident means of distinguishing the account implicit in Figure 2 from that implicit in Figure 3.

Parsons was unhappy with a reductive account of action wherein it was determined by individual ends. He could simply have asserted freedom of will here and denied the predictability of action outright. But instead he systematically linked action to two antecedents: individual ends and social norms. This was because, however unhappy he might have been with reduction, he wanted to retain predictability as a feature of action. Parsons wanted predictability, but he wanted choice as well. And to maintain choice he denied that ends and norms, the predictive factors in this scheme, were causal factors the resultant of which would determine action. Instead, he gave an account (functionally equivalent) of ends and norms as factors in relation to which actions are chosen. But no fact of the matter will allow a decision between this voluntaristic account and a causal one. And indeed the many critics of Parsons who have read his account as a causal one are correct at least in this: there is no sociologically interesting difference between his voluntarism and a causal equivalent such as that in Figure 2.

The tension between ends and norms in inspiring action may be represented either voluntaristically or causally without implications for Parsons's basic sociological purposes. Nor is an understanding of deviance or social change or conflict or any of the other phenomena commonly held to cause difficulties for Parsons affected by which alternative is selected. Why then was Parsons so insistent on voluntarism? The only safe reply is that his insistence was not moved by narrowly technical sociological considerations.

#### AGENCY IN GIDDENS

Anthony Giddens is one of the many critics who argue that Parsons's voluntaristic theory of action is in truth merely a version of determinism. In the development of his theory of structuration (1976, 1979, 1984), Giddens attempts to transcend what he sees as the deter-

minism in Parsons's work on action and insists upon "the freedom of the acting subject." However, Giddens at the same time shares many of the assumptions of the Parsonian approach. Thus, both theorists write against the alleged hegemony of "positivism" and both subscribe to a rarefied Freudianism. Moreover, it is clear that Giddens shares much of Parsons's account of action, intentionality, and meaning, even though he criticizes many of its specific details. An apparently important difference between the two theorists is that Giddens rejects Parsons's division of the "mental" realm into an ego that acts and other compartments of the mind that affect the ego. Likewise, he rejects a division between the mind (including the ego, the conscience, and so forth) and the body. Initially, in Giddens, the actor is an embodied unit and as such, a possessor of causal powers that she may choose to employ to intervene (or not) into the ongoing sequence of events in the world. This makes her an agent.

I shall define action or agency as the stream of actual or contemplated causal interventions of corporeal beings in the ongoing process of events-in-the-world. (1976, 75)

Furthermore, and this is crucial to Giddens's whole argument,

It is analytical to the concept of agency that a person (i.e. an agent) "could have acted otherwise." (1976, 75)

This conception of the agent ties agency to power.

What is the nature of the logical connection between action and power? . . . To be able to "act otherwise" means being able to intervene in the world, or to refrain from such intervention, with the effect of influencing a specific process or state of affairs. This presumes that to be an agent means to be able to deploy (chronically, in the flow of daily life) a range of causal powers, including that of influencing those deployed by others. Action depends upon the capability of the individual to "make a difference" to a pre-existing state of affairs or course of events. An agent ceases to be such if he or she loses the capability to "make a difference," that is, to exercise some sort of power. (1984, 14)

It is worth noting here how Giddens characterizes this power of agents to intervene as a transformative capacity. This suggests that the power to intervene amounts to a power to bring about social change or transformation. But of course the power might be equally well used to intervene in a situation that otherwise would change to

maintain it. What Giddens calls “transformative capacity” could equally well be called “stabilizing capacity,” and it is an interesting reflection on Giddens’s work that it should obscure this point and presuppose a connection between activity and change and, correspondingly, between passivity and stability. For present purposes, however, all that matters in Giddens’s account is his identification of agents as possessors of capacities with which they can choose to intervene.

For Giddens, an agent ceases to be such if she loses the ability to choose. It is precisely according to this criterion of being able to “act otherwise” (and thereby make a difference) that Giddens distinguishes humans from nature. All humans (including social scientists) are active agents, and society is their achievement, just as the ethnomethodologists claim that

the social world, unlike the world of nature, has to be grasped as a skilled accomplishment of active human subjects. (1976, 155)

And, following the ethnomethodologists again, Giddens criticizes Parsons for failing to recognize this:

The use of the term “voluntarism” suggests that Parsons wished to try and build into his own approach a conception of the actor as a creative, innovative agent, thus seeking to break with schemes in which human conduct is not conceptually differentiated from the explanation of the movement of objects in nature. For Parsons the very same values that compose the consensus universal, as “introjected” by actors, are the motivating elements of personality. If these are the “same” values, however, *what leverage can there possibly be for the creative character of human action as nominally presupposed by the term “voluntarism”?* Parsons interprets the latter concept as referring simply to “elements of a normative character”; the “freedom of the acting subject” then becomes reduced—and very clearly so in Parsons’ mature theory—to the need-dispositions of personality. In the “action frame of reference,” “action” itself enters the picture only within the context of an emphasis that sociological accounts of conduct need to be complemented with psychological accounts of “the mechanisms of personality”; *the system is a wholly deterministic one.* Just as *there is no room here for the creative capacity of the subject on the level of the actor,* so there is a major source of difficulty in explaining the origins of transformations of institutionalised value-standards. (1976, 95-96; italics added)

Needless to say, this is a false picture of Parsons’s scheme: it makes no mention of effort; nor does Giddens pay heed to the obvious gap between acting in one of the many ways that conform to a norm and

being completely determined in what one does when conforming to it; nor is there any attention to the differences (represented above as between Figures 2 and 3) between causation and the conditioning of choice on which Parsons lays so much stress. It is useful to attend to this passage, nonetheless, for the insight it provides into the nature of what is in effect Giddens's own voluntarism. Giddens wants people to have choice because he wants them to be capable of effecting change (in the existing order of things). He apparently thinks that only a theory that is voluntaristic in this sense will permit them this capacity and indeed that only a voluntaristic theory of a very strong and comprehensive kind, transcending that of Parsons, will do so. Why he should take this view remains obscure.

Following the ethnomethodologists, Giddens allows his agents discretion over what in Parsons press upon them (and, as Giddens sees it, determine what they do): agents have discretion re norms and rules. But in thus asserting the freedom of agents both from direct determination by rules/norms and from the guilt feelings re rules/norms that in Parsons constrain and press upon choice, Giddens casts aside all the predictive/explanatory features of Parsons's theory. And while this would be unexceptionable for someone who wished simply to proclaim the mystery of free will and the lack of pattern in human actions, it will not suffice for Giddens, who recognizes that routine rule following is indeed very much the most common form of human action and recognizes as well the need to give an account of why this is so.

Here Giddens faces just the same problem as Parsons had faced earlier: how to reconcile choice with pattern and predictability in human action. And his solution is, formally, of just the same kind as Parsons's. The question is: What induces routine norm-conforming behavior, if norm-induced guilt does not? And Giddens answers it by citing a surrogate for guilt, an alternative individual psychic element capable of pressing upon choice—*ontological security*.

For all that his agent is initially an embodied unity, Giddens, like Parsons, comes to recognize a structured human psyche wherein the consciously acting agent is differentiated from an unconscious realm of repressed feelings and uncognizable knowledge. It is here that the motivational basis of action exists, and where the actor's "basic security system" resides, a system that is formed in the early years of life. Because of the existence of this basic security system, actors aim to maintain a high degree of ontological security and hence to sustain

routine at the expense of “disruptive change” that can lead to tension and anxiety.

Actors wants remain rooted in a basic security system, largely unconscious and established in the first years of life. The initial formation of the basic security system may be regarded as involving modes of tension management, in the course of which the child becomes “projected outwards” into the social world, and the foundations of ego-identity created. It seems plausible to suggest that these deep-lying modes of tension management (principally reduction and control of anxiety) are most effective when an individual experiences what Laing calls ontological security. . . . Ontological security can be taken to depend upon the implicit faith actors have in the conventions (codes of signification and forms of normative regulation) via which, in the duality of structure, the reproduction of social life is effected. In most circumstances in social life, the sense of ontological security is routinely grounded in mutual knowledge employed such that interaction is “unproblematic,” or can be largely “taken for granted.” (Giddons 1979, 218-19)

Just as norms press on choice in Parsons, so ontological security presses upon choice in Giddens. Just as an action may be understood as the result of a choice to avoid the pain of guilt in Parsons, so it may be understood as a choice to avoid the anxiety of ontological insecurity in Giddens. And just as a direct causal impact of norms is an alternative formulation to Parsons, so a direct causal impact of the need for security is an alternative to Giddens’s account. One might ask of Giddens, as he does of Parsons, whether his sociology is not “complemented with psychological accounts of ‘the mechanisms of personality.’ ”

In his later work on structuration, Giddens continues in the same vein of identifying constraints on choice (agency). Thus, in *The Constitution of Society* (1984), he highlights three kinds of constraints: material constraints, those deriving from sanctions, and structural constraints. For Giddens, structural constraints both limit possibilities for activity and the generation of outcomes vis-à-vis individuals, and also appear to the agent as prestructured enablements associated with the opportunities for action that remain open. The implication of this is that the idea of structural constraints only makes sense when an individual’s regard for prestructured options is taken into account; that is to say, structural constraints for Giddens are partially constituted by an actor’s motives, wants, and needs. In regard to the latter, Giddens writes,

It is of the first importance to recognise that circumstances of social constraint in which individuals "have no choice" are not to be equated with the dissolution of action as such. To "have no choice" does not mean that action has been replaced by reaction (in the way in which a person blinks when a rapid movement is made near the eyes). This might appear so obvious as not to need saying. But some very prominent schools of social theory, associated mainly with objectivism and with "structural sociology," have not acknowledged this distinction. They have supposed that constraints operate like forces in nature, as if to "have no choice" were equivalent to being driven irresistably and uncomprehendingly by mechanical pressures. Even the threat of death carries no weight unless it is the case that the individual so threatened in some way values life. To say that an individual "had no choice but to act in such and such a way," in a situation of this sort evidently means "Given his/her desire not to die, the only alternative open was to act in the way he or she did." (1984, 175)

And this is the basis upon which he interprets references to structured constraint in the work of the seminal social theorists. Thus,

Marx says that workers "must sell themselves"—or, more accurately, their labour power—to employers. The "must" in the phrase expresses a constraint which derives from the institutional order of modern capitalist enterprise that the worker faces. There is only one course of action open to the worker who has been rendered propertyless—to sell his or her labour power to the capitalist. That is to say, there is only one feasible option, given that the worker has the motivation to wish to survive. (1984, 177)

In summary, Giddens ends up with a theory formally identical to Parsons's voluntarism. It starts with an insistence on choice, and then, in order to allow for patterns and predictability in action, introduces constraints on choice—factors that make it intelligible that choices will be of one kind rather than another. The account of constrained choice in Parsons was indistinguishable in its empirical implications from an account of directly (causally) constrained action (see Figures 2 and 3). The same kind of alternative formulation can be given of Giddens: that needs, and notably the need for ontological security, and similarly constraints, and notably the constraints of structure, may operate upon human beings in a causal sense, with action being understood as following on as the effect of the overall impact of all the various causes. Indeed, it could even be claimed that no formal difference is evident here between Giddens (or Parsons) and mundane rational choice theory, which quite explicitly uses the language of

choice to describe human actions as in principle highly predictable and allows stable wants, desires, or needs of any kind, conscious or unconscious, to be cited in order to make it so. It is useful to recall how rational choice theorists impute agency (choice) precisely to engender accounts of action as predictable/controllable, given that Giddens regards the very same imputation as the necessary means for allowing creativity/innovation.

Giddens himself speaks of actions that "could be otherwise" in order to stress how actions are never wholly determined by structural constraints, but few sociologists have ever believed in complete determination of this kind in any case, and it is readily opposed by the view that if there are structural constraints, they merely feature among the many necessary causes of action rather than counting as sufficient causes of it. There is no need to resort to voluntarism or to insist on the existence of choice or agency in order to set limits on the pretensions of structural explanation. Again, there are times where Giddens adopts a rationalistic approach close to that of Habermas and treats reasons as having the potential to inspire "creative and innovative actions" that depart from routine. But it is perfectly possible to think of being given a reason to act as a causal intervention: if an agent is told there is cyanide in the water and does not drink thereof, it might be said that the proffered reason made her choose not to drink or alternatively that the proffered reason caused her not to, and again how far the agent "could have done otherwise" in the given circumstances, having been told about the cyanide, is unclear and not amenable of empirical investigation.

It is likely that both Parsons and Giddens espouse choice/agency and oppose determinism with much the same kind of end in view, one that is not sociological in a narrow sense. Parsons seeks an actor capable of struggling against self-interest and animal drives. Accordingly, he locates drives and interests within a separate part of the psyche as pressures on the acting ego and gives that ego choice and hence the ability to resist the pressures. Then, in order to make sense of action as patterned, Parsons postulates a second pressure, the pressure of norms, of the internalized moral order, when he is able to speak of choices occurring with frequencies related to the relative magnitude of the two opposed pressures. Analogously, Giddens needs an actor capable of struggling against the status quo and its constituent routines. Accordingly, he internalizes these as a need for ontological security and is able to imagine an agent choosing under the pressure from this need but also under pressures of other kinds. The intention,



no doubt, is to produce a sociologically realistic yet politically optimistic picture of the human condition. But evidence for such a picture, and in particular for the role of agency within it, is not supplied.

There is no need for assertions of choice and/or agency here in order to be sociologically realistic. It is perfectly possible to produce empirically plausible accounts of the relationship of actions to self-interest or to the social status quo without using such vocabulary. Nor is this vocabulary especially appropriate to the expression of political optimism and the conviction that human beings may act in ways that overcome external pressures and restraints. A voluntaristic style of discourse may have suited the libertarian socialism of Giddens, but it has also suited the objectives of repressive political and religious regimes, which have sought to constrain their subjects precisely by stressing their freedom of action and making them responsible and accountable for what they do—with their lives in some cases. Conversely, fully causal accounts of action, for example, those in the various theological doctrines of predestination and divine determination, have been adopted by collectives concerned precisely to ignore and contravene the authority of church and state and even actively to oppose and overthrow them. Oddly perhaps, but oddly only to us, through and beyond the Reformation, it suited creative and resourceful opponents of the political and institutional status quo to hold that, of themselves, they could not have acted otherwise.

#### SOCIOLOGY AND THE INSTITUTION OF RESPONSIBLE ACTION

When a human being acts, it may be regarded as the implementation of a choice or as the effect of a cause or causes. In either case, the conditions and circumstances of what is done will be relevant to understanding what occurs: they will be taken account of as a choice is made, or else they will feature as the necessary conditions in which a cause will bring about its specific effect. It is sometimes said that if we speak of causation should we speak of behavior, and that only if we speak of choice we should speak correspondingly of action, but this is a dubious claim. Action in the sense of meaningful behavior may be held nonetheless to be caused without obvious difficulty. Suicide “caused by” clinical depression need not be regarded as “mere behavior” rather than as meaningful action, nor need recognition of a

suicide as meaningful action preclude consideration of whether clinical depression might have caused it.

Whether an action is chosen or caused has been a significant question not just for sociologists but for ordinary members of society as they have sought to make sense of how other people are acting. Any action may be rendered according to elements of the institution of responsible action or according to elements of the institution of causal connection. The apparent difficulty is that while there are clear and evident patterns in how members deploy the resources represented by these two institutions, there is no fact of the matter that serves to make sense of these patterns. It is not possible to examine the antecedents of actions and find a feature of caused actions that is not possessed by chosen ones. Action that "could have been otherwise" is not obviously different in its characteristics or its ancestry from action that "could not have been otherwise."

A standard approach to a theoretical difficulty of this kind would be to adopt the perspectives of ethnomethodology and to speak of members achieving a shared sense of an action being caused (or chosen) as their collective accomplishment. Imagine that the action itself, as it were, will accede to either account indifferently and that the actual account given to it is purely a matter of what members are disposed to give it. Extension of this argument could lead to the thought that sociological theorists are voluntarists or determinists according to taste, that the issue of which alternative is preferable is undecidable by any standard sociological argument, and that the only sociologically interesting question here might be what it is about theorists that makes sense of their preferences for the one or the other.

There is indeed much to be said for an approach of just this kind, and it may well be that it provides the only kind of account that will always be relevant to ascriptions of choice or causation. There is only one past. Whether or not it could have been otherwise, it was not otherwise, and nothing empirical hangs on the might have been that was not. Theorists may ascribe causal or noncausal antecedents to past actions, safe in the knowledge that nothing will refute their favored kind of imputation. Nor will it be material whether predictions of future action are based on the supposition that relevant factors operate as direct causes or on the alternative view that sees them as influences on choice. Indeed, it is even possible to find hybrid formulations in social theory, such as those that ask what factors determine members' choices, and even philosophically odd approaches such as these are serviceable for many purposes.

Nonetheless, we should not assume, because of what has been said so far, that accounts of action as caused or chosen invariably have nothing to do with the action and everything to do with those who account it the one or the other. That these accounts are not efforts to describe the manifest antecedents of action does not mean that they are altogether lacking in any connection with observable features of courses of action and the situations wherein they occur. Indeed, there is a very obvious and widespread pattern in the mundane use of the institution of responsible action, and the contrasted use of the institution of causal connection, which suggests that such observable features may very well be at work in their deployment. Although the institution of responsible action is capable of many and various uses, and its "correct use" is always in the last analysis a contested matter on which sociologists have no standing to pronounce, it is worthwhile nonetheless to look at what is currently one of the paradigmatic patterns of use of the institution and what informs it.

If we say that an action was chosen and "could have been otherwise," what does that signify? The immediate thought is that an absence of causation is indicated, the existence of an act of will uncoupled from a causal nexus, perhaps even the intervention of a nonmaterial agency into the ongoing sequence of a material world. But none of this has any empirical significance. Is there then anything at all that does have such significance that could be learned or inferred from an imputation of choice? Consider this possibility: a chosen action, in being accounted chosen, is identified as the kind of action that could be modified or inhibited by symbolic communication, or as we often say, by persuasion. If this conjecture is correct, then there is indeed genuine empirical point in identifying actions as chosen. By this means we make visible kinds of action, of actor, of action-situation, where resort to persuasive intervention might perhaps be worthwhile, where there is some hope that it could result in success. The language of choice and volition now serves to mark out situations wherein the use of persuasion might be worthwhile and hence makes a genuine empirical contribution to the stock of knowledge of competent members, a contribution that facilitates the effective use of persuasive interventions in future situations. Any specific deployment of the discourse of the institution of responsible action, on this account, orients to a theory about when, where, and to what degree there is point in treating others as "persuadable," "amenable to reason"; in short, as moral agents.

An account of this kind does fit reasonably well with many routine commonsense references to choice and agency in the context of the institution of responsible action. In extreme cases of allegedly involuntary, caused, not-chosen behavior, our designations tend precisely to be based upon the existence of action that cannot (we believe) easily be modified, upon "compulsive" or "fixated" patterns of behaviors. Indeed, in these cases, the cases of the severely mentally ill and the criminally insane, it may not be just persuasive communication but all forms of intervention short of use of a straitjacket that fail to modify the patterns. Conversely, at the other extreme, there is a great range of chosen activity that is (we imagine) extremely easy to modify by intervention: there is the walk that may be modified to include a call at the shop (Would you mind?), the cigarette that may be extinguished on request (Did you not notice the sign over there?), and endless similar cases. And there are the inevitable problematic cases between the two extremes, such as those actions accompanied by apology—"I'm sorry, I had no choice"—which typically means "I did have a choice, but only at some considerable cost." With actions of this last kind, only strong intervention is going to modify what is done, and it may be that no merely persuasive communication or exhortation will serve to divert the actor from doing it.

What we have here is a range of courses of action from those that are extremely difficult to modify, through to those that may be modified by the most cursory intervention—a mere word or two sufficing. Nowhere in the range is it at all material whether the courses of action really are caused, whether they "could not have been otherwise." From the point of view of a causal analysis, a course of action that is ongoing in a "cannot be otherwise" way, because of the operation of a set of constantly operating causes, may nonetheless be modified if an additional cause impinges upon it—and intervention with persuasive communication just is such an additional cause, a disturbance, which, if it leads to variation in the ongoing course of action, leads to variation that itself "could not have been otherwise." Thus, from the point of view of a believer in universal causation the everyday contrast of chosen and caused behavior does not refer to the presence or absence of causation but rather to the degree of resistance to change of a (caused) course of action. Consider how a leaf falls from a tree and how a branch falls: both are conventionally accounted as caused movements, but the slightest breath of wind will vary the path of the leaf while leaving the branch moving as before. It is the difference between the leaf and the branch (a believer in universal causation

might claim) that everyday deployment of the institution of responsible action effectively stresses.

Nor need a believer in the ubiquitous existence of choice and free will wholly reject this kind of analysis. Might it not be that all actions are chosen but that there is a range of chosen actions from those readily modified to those carried out with implacable will and determination? Might not normal practice merely identify these two extremes of choice as voluntary and caused so that the caused actions of the mentally ill and insane are merely actions characterized by an implacable willfulness?

The metaphysics of the ordinary member in this context may be criticized from the standpoint of believers in universal causation or that of believers in universal human freedom. But the associated discourse of the ordinary member nonetheless does an empirically valuable job of work, using the institutions of responsible action and causal connection to map the susceptibility of actors to persuasion. Such discourse is an important backdrop to the use of argumentation and moral evaluation in the modification of the actions of other people.

Consider now the legitimate interests of social theorists when they address the discourse and the action of ordinary members. Theorists need to refer to the antecedents of action to make that action intelligible and predictable. For this purpose, it is immaterial whether they ask how the antecedents feature among the causes of action, or how they condition the choice of a course of action. Theorists also need to know how resistant to modification different actions may be. For this purpose, again, it is immaterial whether an action is taken to be caused or chosen. Theorists do not need to know, however, how far a course of action "could have been otherwise," for there is nothing here to be known. And it is only for this unrealizable and misconceived purpose that it might be said to matter whether an action is "really" caused or chosen. Thus, in speaking of action and its antecedents, it makes no practical difference whether the language of causation is employed or that of choice and agency. Indeed, it is perhaps because it makes no difference one way or the other in this sense that ordinary members put the language of agency and causation to another use and tend to the custom of referring to causation to index how resistant to modification through communicative interaction an action is considered to be.

In social theory, the choice between voluntarism and causation is widely held to be of great import, yet detailed analysis reveals the

irrelevance of the issue to sociological concerns. Parsons makes an immense effort to formulate his theory as a voluntaristic one, and it cannot be said that he fails to do so. But Giddens and other critics nonetheless perceive the theory as causal and even completely deterministic. This is not because the critics are straightforwardly wrong, although they are surely perversely insensitive and unsympathetic in their treatment of Parsons, and no more right than he is. It is because the voluntarism Parsons developed with so much effort is, from a sociological perspective, no more than a surface gloss. Sociological theories need no protection from the weather, and there is nothing for the surface gloss to do, certainly nothing that a wholly causal gloss would not do just as well. This no doubt is why Giddens in his turn has found his own theory read as a deterministic one despite its emphasis on human agency (Thompson 1989, 73-74). Insofar as the understanding of voluntary action is concerned, Giddens's encounter with the problem, despite first appearances, has largely been a recapitulation of Parsons's encounter. Insofar as sociology is concerned, neither encounter and neither way of responding to it is of interest: indeed, the responses of both Giddens and Parsons reflect considerations that go way beyond the proper bounds of sociology. To recognize this is to recognize agency as a red herring in the context of sociological theory.

#### REFERENCES

- Barnes, B. 2000. *Understanding agency: Social theory and responsible action*. Thousand Oaks, CA: Sage.
- Giddens, A. 1976. *New rules of sociological method*. London: Hutchinson.
- . 1979. *Central problems in social theory*. London: Macmillan.
- . 1984. *The constitution of society*. Cambridge, MA: Polity.
- Loyal, S. Forthcoming. *The sociology of Anthony Giddens*. London: Pluto.
- Parsons, T. 1937. *The structure of social action*. New York: Free Press.
- Thompson, J. 1989. The theory of structuration. In *Social theory of modern society: Anthony Giddens and his critics*, edited by D. Held and J. Thompson. Cambridge, MA: Polity.

*Steven Loyal teaches in the Department of Sociology, University College Dublin, Ireland.*

*Barry Barnes teaches in the Department of Sociology, University of Exeter, England.*

## Discussions

---

### Sexual Harassment and Wrongful Communication

EDMUND WALL  
*East Carolina University*

In the first part of this discussion, I argue that we need a new model of sexual harassment. A case is made that the prevailing model, based on sex discrimination and endorsed by feminists such as Catharine MacKinnon, is incompatible with fundamental democratic values. In the second part of this discussion, I defend an account of sexual harassment modeled on disrespectful communication.

#### I

MacKinnon and other feminists argue that sexuality and gender are intimately linked. MacKinnon (1987, 6) maintains that gender is a “congealed form of the sexualization of inequality between men and women.” According to MacKinnon (1997, 3, 37, 39), gender is not a difference in nature but a hierarchy of power in which women are dominated by men.<sup>1</sup> We are also told that the abusive treatment of women by men, which, among other things, includes rape, assault, and harassment, is a form of sex for men. Since, in MacKinnon’s estimation, whatever is felt as sexual is sexual, such acts of dominance, submission, and violence are all aspects of sex. Men—and women—are said to be sexually aroused by such actions. Given that in our society a violation of the powerless is sexy, and violation of the

---

This discussion benefited considerably from the thoughtful suggestions of Burleigh Wilkins, who read an earlier version of it. The editorial revisions suggested were very helpful as well.

*Philosophy of the Social Sciences*, Vol. 31 No. 4, December 2001 525-537  
© 2001 Sage Publications

powerless is central to the very meaning of female and male, gender should not be considered apart from sex and dominance in MacKinnon's (1987, 6) estimation.

These background assumptions can explain why MacKinnon and similar feminists call for a very broad definition of sex discrimination, a definition that runs the gamut of sexual issues. When MacKinnon claims that the harassment of women is based on sex, the term *sex* includes a combination of sexuality, power, and gender. Her approach purports to accurately depict social reality rather than to construct a systematic set of abstract principles (MacKinnon 1987, 40-41; see also Feary 1994, 649-62).<sup>2</sup> All of this may explain why MacKinnon and other feminists talk about sex discrimination in such broad terms, but it does not justify their doing so. If we follow their lead and define gender broadly, we cannot offer an adequate explanation of a bisexual harasser who victimizes both women and men. Even if women have been socially constructed to be the sexual slaves of men, the women harassed by the bisexual perpetrator are not harassed because they are women any more than the male victims of the bisexual perpetrator are harassed because they are men. In such cases, the issue is not gender but rather a disregard for basic human dignity.

MacKinnon has argued that the question whether a given case of harassment is based on sex or personal considerations rests on a false dichotomy. She believes that to relegate sexual harassment cases to the category of a "personal episode" is to isolate and further subjugate the victims by stigmatizing them as deviants. In MacKinnon's estimation, the harassment of women is based on sex. It is done because they are women. Here, MacKinnon (1987, 106-7) argues that a woman is not only a woman personally but also socially. Membership in a gender is a part of a woman's individuality.

It would be hard to see how anyone could consider the sexual harassment of female and male employees to be a "personal episode," but let us consider MacKinnon's point that a woman's experiences and very identity are inseparable from her social status and that, therefore, her victimization implies gender discrimination. The response is that the bisexual harasser does not single out women as a group, or men as a group for that matter. The perpetrator's lack of respect is indicative of what he has let himself become and not indicative of some point of view about women per se. Moreover, there are many individuals in our society who treat all other human beings with respect, regardless of their gender. Indeed, to say that someone treats human beings of one gender with respect but not those of the



other gender does not ring true. We would seriously question whether such an individual could treat any human being with genuine respect. This concern would arise precisely because all human beings are alike in some fundamental ways. They want to be considered important and not merely in an instrumental way. They want the type of thoughtful treatment that is compatible with their sense of dignity and intrinsic worth. These common threads that characterize all healthy human beings are obfuscated by MacKinnon's emphasis on the male dominance of women. Her broad definition of sex discrimination, as tempting as it may be to some feminists, may not advance her attempt to characterize social reality.

MacKinnon is not merely offering an interesting account of sexual harassment but a model for use in federal civil rights litigation as well. Indeed, MacKinnon's work on sexual harassment has influenced the direction of sexual harassment law. Title VII of the Congressional Civil Rights Act of 1964 is now being applied to sexual harassment civil suits, and MacKinnon's influence has had something to do with this development. The definition of sexual harassment that is employed in such litigation has been established by the Equal Employment Opportunity Commission (EEOC). Here is the EEOC's definition of sexual harassment:

Unwelcome sexual advances, requests for sexual favors, and other verbal or physical conduct of a sexual nature constitute sexual harassment when (1) submission to such conduct is made either explicitly or implicitly a term or condition of an individual's employment, (2) submission to or rejection of such conduct by an individual is used as the basis for employment decisions affecting such an individual, or (3) such conduct has the purpose or effect of unreasonably interfering with an individual's work performance or creating an intimidating, hostile, or offensive working environment.<sup>3</sup> (Wall 1991, 382)

Before examining this definition, it should be pointed out that the definition of sexual harassment poses an enormous problem for social researchers. Disagreements about the definition as well as differences in survey techniques have prompted many social scientists to throw up their hands at comparisons between studies (see Fitzgerald, Swan, and Magley 1997; Wall 1992).<sup>4</sup> The project of ensuring reliable research data on sexual harassment cannot be separated from the quest for a philosophically accurate definition. To study sexual harassment, we must first be able to identify it. Moreover, the quest for an acceptable legal definition raises profound philosophical ques-

tions pertaining to justice. A democratic society requires a definition that will identify sexual harassment without violating the moral rights of the innocent. The EEOC's definition targets, among other behaviors, "verbal or physical conduct of a sexual nature" that create an "intimidating, hostile, or offensive working environment." That is too broad.

Suppose that an employee installs a nude painting above his office desk. After doing so, let us say that he never makes reference to it. The painting has been added solely for the employee's own benefit. Given the volatility of the sexual harassment issue, we surely can envision strong protests by other employees against the painting. In this case, the employee's "physical conduct of a sexual nature" has created a "hostile" working environment. Indeed, the other employees can say, in good faith, that the painting has created a hostile working environment. Certainly, the painting may be inappropriate in the workplace, and the other employees may be justified in requesting that it be removed (although even this conclusion is not beyond contention), but should our legal net capture cases such as this? Is the art-loving employee a sexual harasser? Such a judgment seems to be unjustified. In constructing a definition of sexual harassment, one ought to be concerned with the requirements of justice. Among other things, a democratic society aspires to protect its citizens against the massive power of the federal government.

Not only does the EEOC's definition capture cases that may not involve sexual harassment, it also fails to capture obvious cases of sexual harassment. Suppose a female employee is constantly being stared at in a suggestive way by a male employee. Although this behavior may upset the employee, let us assume that she decides to shrug it off as one of the many drawbacks of the workplace. Let us say that after many long years of enduring similar behavior she has built up psychological defenses against it. Certainly MacKinnon would agree that this must be sexual harassment and that it can be just as objectionable as a case involving verbal harassment. Yet the EEOC's definition fails to capture cases such as this. After all, "unwelcome sexual advances" or "physical conduct of a sexual nature" are to be labeled as sexual harassment only when one of three conditions are met. We have seen that the first two conditions require more overt behavior (i.e., *quid pro quo* sexual harassment) and therefore would not apply to this case. The third condition that seeks to identify hostile environment cases would seem to apply but comes up short. The unwelcome sexual advances may not have interfered with the

employee's work performance, and since she and perhaps the other employees have built up defenses against the inappropriate advances, have not created an "intimidating, hostile, or offensive working environment." Thus, it seems that the victimized employee would not have a legal case against the harasser.

Feminists such as MacKinnon insist that, in order to be sexually harassed, a victim need not actually experience any emotional distress. They endorse a reasonable person/victim standard (MacKinnon 1992a; Rigor 1992; Superson 1993; Ehrenreich 1992).<sup>5</sup> The proper inquiry, they say, is whether a reasonable person/victim would have felt distressed by the behavior in question. They argue that to focus on the feelings of the actual victim fails to acknowledge that women have been conditioned to accept a great deal of harmful sexual behavior (MacKinnon 1992b; Feary 1992; Superson 1993). They approvingly point out that the EEOC does not require the actual victim to experience emotional distress.

There is a serious issue here pertaining to individual freedom and autonomy. Sexual harassment is a type of harassment. When an alleged victim is not actually bothered by the behavior in question, we cannot rightfully say that she has been harassed. To design a legal system so that it offers protection against harassment to people who may not experience any distress (indeed, who may sincerely believe they have not been harassed) is antidemocratic. The EEOC's definition runs roughshod over basic moral rights. And just who are the beneficiaries of this legal paternalism? MacKinnon and other feminists have made a case that women have been conditioned throughout the years to accept inappropriate sexual advances. Since they have not made the case that men have been similarly conditioned, they have not provided sufficient assurance that men, also, should be able to claim sexual harassment in the absence of any emotional distress. But basic justice requires such an assurance.

Instead of concentrating on issues of gender and discrimination, we would do well to consider how human beings communicate with one another. Looking at sports events, political debates, popular media ads, and other aspects of our everyday interactions, we find disrespect. Consider that all too often we walk down a street and a stranger will fix a menacing stare on us. This is not a crime. It may not be a violation of our moral rights, either. Nevertheless, these gestures are anything but benign. They communicate disrespect for human beings. If communication is to be understood broadly, and if disrespectful communication has become commonplace, it is no wonder

that sexual harassment is such a pervasive problem. If it is an issue of power and control, then the discussion of it should encompass the pervasive social maladies just described. Low self-esteem, cynicism, and hatred of self and others should be a major focus of the discussion. It is also true, however, that to help ensure basic justice, we ought to implement legal remedies fairly. It is, therefore, necessary to distinguish sexual harassment from other disrespectful actions. Unfortunately, since we live in a society in which disrespect is pervasive, it is doubtful that any legal remedy will make a significant difference.

The model of sexual harassment as sex discrimination offers harmful generalizations about women and men. Generally speaking, women have not been conditioned to accept inappropriate sexual advances. Philosophers have correctly denounced this assumption as patronizing (Paul 1992). Neither do men, generally speaking, make a practice of demeaning women. Both assumptions lead to a dead end, with a lot of ill feeling along the way. Rather than encouraging respect and healthy relationships between women and men, *these assumptions are actually symptoms of the main problem*. These are demeaning assumptions about women and men, which is another way of saying that they convey a disrespect for human beings.

There is no adequate justification for the view that sexual harassment is a form of sex discrimination. A model of sexual harassment that centers on disrespectful communication seems to be a step in the right direction. Whether it is a menacing stare, slanted remarks about others, the harassment of an employee, and so on, the root of the problem is what we think and feel about ourselves.

## II<sup>6</sup>

I begin my definition of sexual harassment with the following condition: X communicates to Y, X's or someone else's purported sexual interest in Y, or interest in someone else. It seems that the alleged sexual interest need not be in the recipient of the communication for the discussion of that alleged interest to harass the recipient. The discussion might concern an alleged interest in some third party and be every bit as objectionable to Y as a discussion of an interest in Y herself. "Sexual interest" is meant to capture not only sexual attraction but also curiosity about another person's sexual behavior or thought processes. The presence of other objectives such as a desire to dominate another person's sexual behavior or thought processes may indi-

cate that X does not have a genuine sexual interest in Y or some third party, and thus I make room for this by referring to the “purported” sexual interest in Y. If these other objectives are coupled with the above-mentioned curiosity or with a sexual attraction, which is typical in these cases, then the harasser would have a sexual interest in Y. In any case, the term *sexual interest* is deliberately meant to be broad. Harassers may or may not be sexually attracted to their victims. Thus, on my use of the term, a sexual interest in Y does not imply a sexual attraction to Y, although it certainly could amount to that.

*Harassment* suggests not only that someone has been emotionally upset by someone else’s sexual advances but that the sexual advances are repeated over and against the victim’s disapproval. A harassment victim is certainly bothered by a perpetrator’s advances, but *harassment* suggests that a victim is actually hounded by a perpetrator. A deplorable, one-time sexual advance may violate a victim’s moral rights, but it is to be distinguished from repeated and deplorable sexual advances which, due to the repetition, could constitute cases of harassment. Depending on a perpetrator’s intentions, the nature of the action, the circumstances, and so forth, a one-time sexual advance may be more morally repugnant than a series of unacceptable sexual advances, but the one-time offense does not constitute harassment. It may be a full-blown sexual assault, for example, and as such, it should be legally actionable. But the paradigm of sexual harassment involves a series of less egregious (though disrespectful) sexual advances.

In light of these considerations, I propose the following set of necessary and jointly sufficient conditions of sexual harassment:

1. X successfully communicates to Y, X’s or someone else’s purported sexual interest in someone (whether Y or someone else).
2. Y does not consent to discuss or consider such a message about X’s or someone else’s purported sexual interest in someone.
3. Disregarding the absence of Y’s consent, X repeats a message of this form to Y.
4. Y feels emotionally distressed because of X’s disregard for the absence of Y’s consent to discuss or consider such a message and/or because Y objects to the content of X’s sexual comments.<sup>7</sup> (Wall 1991)

When sexual harassment occurs, the perpetrator has engaged in disrespectful communication with the victim. He has successfully communicated some sexual message to the victim, which the victim does not choose to discuss or to consider, and which the perpetrator

continues to convey to the unwilling victim. The victim, in turn, is bothered by the repeated advances and thus the harassment.

The second condition requires that the victim does not consent to receive the sexual message put forward by the perpetrator. This lack of consent can take more than one form. The victim may verbally, or by gesturing, convey an objection to the perpetrator in response to the advance, or the victim may maintain a suggestive silence in response to it. Remember, at this point, a perpetrator would have successfully conveyed a sexual message to the victim. If, for example, the perpetrator has not conveyed the message (for example, perhaps the victim did not hear the perpetrator), or if he did convey the message but, for some good reason, believes that he did not do so successfully, then genuine communication has not occurred. The perpetrator has not yet harassed the victim. He is entitled to determine whether the victim has received the message. But when the victim has received the message, and the perpetrator has good reason to believe that the victim has received the message, then repeating the message can constitute sexual harassment. Sexual harassment can be either intentional or the result of negligence. With regard to the latter, sometimes an insensitive individual will not realize that he is harassing others, but he is still morally responsible for his disregard of the moral rights of others. The fundamental point here is that the victim does not consent to the communication. In the absence of some overriding moral commitments, people ought to be morally sensitive to what others choose to discuss.

It might be said that the proposed conditions could characterize an innocent individual as a harasser. For example, suppose that a member of one culture makes remarks of a sexual nature to a member of another culture, not realizing that the latter would be offended by those remarks. It seems in this case that the so-called perpetrator acted reasonably, given his cultural indoctrination and the information that he had at the time. Is this a genuine counterexample to the proposed conditions? On further reflection, it would seem otherwise. The proposed conditions require that the sexual advances continue even though the recipient of the advances either expresses disapproval or chooses not to take part in the communication. Even if there exist cultures in which individuals are expected to engage in sexual communications of which they disapprove and of which they choose not to engage, it still would seem that this particular practice could be called into question on the grounds of basic respect and autonomy. If a social practice is occurring, that is not adequate reason (or even a reason) for concluding that such a practice should occur. Even if disre-

gard for another individual's autonomous choices is prevalent within a culture, we may still criticize such an attitude on moral grounds.

Another concern may arise, this time with regard to the third condition, which requires that the sexual advances be repeated. One might object that a single, one-hour sexual advance would seem adequate for sexual harassment to occur. It also would seem that the proposed conditions cannot capture this point. But notice that an hour-long episode actually would involve a pattern of sexual advances and not merely one such advance. We are not talking here about one sexual advance made within a 60-minute time frame but rather a nauseating array of sexual advances spread over a duration of 60 minutes.

Yet another line of criticism is that if Y is genuinely indifferent to X's advances and thereby does not consent to the sexual advances, she might not be harassed, even though the proposed account would seem to indicate otherwise. As the objection goes, she would not be harassed because she would not have expressed dissent to the sexual advances. My response is that such indifference would indicate that the fourth condition would not have been satisfied. The fourth condition requires that Y feel emotionally distressed by X's sexual advances. When each of the four conditions are satisfied, Y feels emotionally distressed about repeated sexual advances to which she does not consent. Thus, the merely indifferent nonvictim does not pose a problem for the proposed account of sexual harassment.

What about the nature of the moral rights of these victims? How does sexual harassment infringe on those rights, and what rights are specifically violated by perpetrators? These rights are to be identified as privacy rights and autonomy rights. More attention should be given to autonomy rights, however, as they are more fundamental. As Leeser and O'Donohue (1997) correctly point out in "Normative Issues in Defining Sexual Harassment," an adequate account of privacy rights will not equate such rights with a right to be shielded from all disturbances.<sup>8</sup> It will not suffice to say that if a perpetrator fails to gain the victim's consent to the communication, the sexual advance constitutes a privacy rights violation. Leeser and O'Donohue are correctly concerned that such a broad account of privacy rights may infringe on the rights of free expression. It is, indeed, a mistake to say that any and all sexual advances to which a victim is an unwilling party constitute an infringement on the victim's privacy rights. Moreover, as the authors suggest, one ought to distinguish between a failure to show proper respect for a victim's privacy rights and the actual



violation of a victim's privacy rights. The violation of such rights would involve the actual intrusion into privileged areas (i.e., illegitimate access to privileged information).

What should be said about moral rights and inappropriate sexual advances? Such advances may not themselves violate an individual's right to privacy, but they are potential encroachments on an individual's right to privacy. Such advances seek to elicit some response from the victim. The victim is being encouraged to respond to some sexual advance or to respond to the pressure exerted by the perpetrator, who has repeatedly communicated a message to the victim without concern for the victim's consent to that communication. Such activity constitutes a potential encroachment on a victim's privacy rights, as the victim is being encouraged to discuss sexual matters, despite her objections to doing so. Whether an actual privacy rights violation occurs is determined by the nature and extent of the pressure exerted by a perpetrator on a victim and also by just what it is that the victim is being asked to do. If access to privileged information is not involved, then privacy rights would not be violated. The victim's autonomy rights would be violated, however.

The fundamental fault with the perpetrator's approach is that he disregards the victim's autonomous choice. The victim does not consent to the sexual advance, and yet the perpetrator persists with more advances. All cases of sexual harassment involve a lack of respect for a victim's autonomy rights. Rather than supporting and promoting autonomous choice, a perpetrator discourages it by showing disrespect for the victim's choice. The perpetrator exhibits more than just disrespect for autonomy rights, however. There is one choice that a victim cannot bring to fruition, one choice that remains ineffectual in the face of the harassment, and that is the choice not to be subjected to the sexual communication. In this way, sexual harassment encroaches on the autonomy of its victims. The victim cannot realize, in the presence of the perpetrator, her reasonable expectations about responsible interpersonal relations. The perpetrator's repeated sexual communication, which imposes obstacles to the realization of the victim's reasonable choices, lies at the root of sexual harassment.

#### NOTES

1. Camille Paglia (1992) strongly disagrees. She believes that gender does reflect a difference in nature. She disagrees with other points that can be attributed to



MacKinnon and other feminists of a similar bent. Ellen Frankel Paul (1992) does not believe that women are victims of their social class and thus are dominated by men. Paglia and Paul may be considered feminists, but their approach to social and political relations differs considerably from feminists of MacKinnon's persuasion.

2. Feary (1994) adopts MacKinnon's basic social assumptions to make some of her own points. She critiques my earlier article on sexual harassment (Wall 1991). I thoroughly disagree with Feary's critique of my article, but perhaps the two most significant points of departure are as follows. Feary argues that the case of a bisexual manager, one who harasses both women and men, would not offer a genuine counterexample to her view that sexual harassment against women constitutes sex discrimination. She argues that, in such cases, "sexual issues" such as "sexual preferences, sexual orientation," and so forth would be used as a basis for employment decisions, and that this would support her point about sex discrimination (Feary 1994, 655). My response is that the debate pertained to sex discrimination against women as a group. The discussion was supposed to be about gender discrimination, not sexual preference or orientation. Certainly, these considerations to which Feary refers would constitute irrelevant bases for employment decisions, but that is not a reason for defining sex discrimination in terms of such sexual issues. When the bisexual manager harasses a heterosexual employee (i.e., an employee who has that sexual orientation or preference), the point is not that the manager would be discriminating against heterosexuals. Gender discrimination is the issue. The other main point of contention between Feary's essay and my earlier essay centers on my fourth condition that requires that sexual harassment result in emotional distress in a victim, because the harasser does not attempt to obtain the victim's consent to the sexual discussion and/or because the victim objects to the content of the harasser's sexual comments (Wall 1991, 374). Feary prefers a reasonable person standard to the subjective standard that I employ (Feary 1994, 655-56). I respond to this objection in the text of this discussion, and I do so by appealing to principles of justice and freedom. But, here, I would like to address Feary's analysis that, on my definition of sexual harassment, "the content of what is communicated is immaterial" (p. 657). This is simply not true. My fourth condition of sexual harassment explicitly refers to the content of the perpetrator's sexual comments. This oversight with regard to my fourth condition prompts Feary to draw some incorrect conclusions. For example, in reference to the Senate confirmation hearings on Clarence Thomas's nomination to the U.S. Supreme Court, Feary says that, assuming the charges of sexual harassment against Judge Clarence Thomas were true, my definition would fail to capture the harassment of Anita Hill. Feary believes that my definition would fail to acknowledge "the display of objectionable sexual objects." But my fourth condition does acknowledge such objectionable displays. Moreover, on my account, the perpetrator need not verbalize a sexual interest in the victim. Neither does my account require that the perpetrator's sexual advance be limited to the sex act. Any communicated interest of a sexual nature would have been captured by the definition that I offered.

3. Feary (1994) defends the Equal Employment Opportunity Commission's (EEOC) definition of sexual harassment and argues that sexual harassment is not a major philosophical problem. In the text of this discussion, I will offer arguments against the EEOC's definition, but let me point out that another feminist who shares Feary's basic assumptions has argued against the EEOC's definition (see Superson 1993).

4. There is a plethora of evidence to suggest that the definition of sexual harassment poses a serious problem to social researchers. For a more recent overview of the scope and differences between definitions, see Fitzgerald, Swan, and Magley (1997). Fitzger-

ald, Swan, and Magley provide an extensive bibliography as well. In Wall (1992, 19, n. 1), there is an ample list of social research in the 1980s and early 1990s on the problem of defining sexual harassment.

5. Nancy S. Ehrenreich (1992) questions the effectiveness of the reasonable woman standard. Although she believes it to be an improvement on the reasonable man standard, she questions whether our society can adequately construct any equitable social standards given the fact that women and men, in her estimation, are the victims of their respective social and economic classes.

6. From Edmund Wall, *Sexual Harassment: Confrontation and Decisions*, pp. 99-111 (Amherst, NY: Prometheus Books). Copyright © 2000. Reprinted by permission of the publisher.

7. These conditions have been revised from the conditions that I offered in Wall (1991). A new analysis is offered here as well.

8. Leeser and O'Donohue (1997, 44) have correctly criticized my previous definition of sexual harassment. On pages 38-43, Leeser and O'Donohue also critique my definition of coercion, but their objections to that definition do not seem convincing. I have constructed a response to their objections in an unpublished manuscript, "A Theory of Coercion."

## REFERENCES

- Ehrenreich, Nancy S. 1992. Pluralist myths and powerless men: The ideology of reasonableness in sexual harassment law. Reprinted in *Sexual harassment: Confrontations and decisions*, edited by Edmund Wall, 229-60. Buffalo, NY: Prometheus Books.
- Feary, Vaughana Macy. 1994. Sexual harassment: Why the corporate world still doesn't "get it." *Journal of Business Ethics* 13:649-62.
- Fitzgerald, Louise F., Suzanne Swan, and Vicki J. Magley. 1997. But was it really sexual harassment? Legal, behavioral, and psychological definitions of the workplace victimization of women. In *Sexual harassment: Theory, research, and treatment*, edited by William O'Donohue, 5-28. Boston: Allyn & Bacon.
- Leeser, Jaimie, and William O'Donohue. 1997. Normative issues in defining sexual harassment. *Sexual harassment: Theory, research, and treatment*, edited by William O'Donohue. Boston: Allyn & Bacon.
- MacKinnon, Catharine A. 1987. *Feminism unmodified: Discourse on life and law*. Cambridge, MA: Harvard University Press.
- . 1992a. Sexual harassment as sex discrimination. Reprinted in *Sexual harassment: Confrontations and decisions*, edited by Edmund Wall, 139-50. Buffalo, NY: Prometheus Books.
- . 1992b. The social causes of sexual harassment. Reprinted in *Sexual harassment: Confrontations and decisions*, edited by Edmund Wall, 121-36. Buffalo, NY: Prometheus Books.
- Paglia, Camille. 1992. Sexual aggression and nature. Reprinted in *Sexual harassment: Confrontations and decisions*, edited by Edmund Wall, 112-13. Buffalo, NY: Prometheus Books.
- Paul, Ellen Frankel. 1992. Bared buttocks and federal cases. Reprinted in *Sexual harassment: Confrontations and decisions*, edited by Edmund Wall, 151-57. Buffalo, NY: Prometheus Books.

- Riger, Stephanie. 1992. Gender dilemmas in sexual harassment policies and procedures. Reprinted in *Sexual harassment: Confrontations and decisions*, edited by Edmund Wall, 197-215. Buffalo, NY: Prometheus Books.
- Superson, Anita M. 1993. A feminist definition of sexual harassment. *Journal of Social Philosophy* 24 (1): 46-64.
- Wall, Edmund. 1991. The definition of sexual harassment. *Public Affairs Quarterly* 5 (4): 371-85.
- , ed. 1992. *Sexual harassment: Confrontations and decisions*. Buffalo, NY: Prometheus Books.

# Rewriting Color

B.A.C. SAUNDERS  
J. VAN BRAKEL  
*K. U. Leuven, Belgium*

Berlin and Kay ([1969] 1991) claimed that in all languages there are words referring to the same 2 to 11 basic color terms (henceforth BCTs). These emerge in a fixed, unilinear evolutionary sequence: WHITE and BLACK, RED, GREEN/yellow or YELLOW/green, blue, brown, purple/pink/orange/gray. Equal probabilities of evolution are indicated by a slash; uppercase terms refer to the early evolutionary stages or composite categories.<sup>1</sup> As some linguistic communities do not lexicalize all 11 BCTs, the question was raised whether the BCTs were already in the head awaiting their evolutionary triggering. The work of Rosch (1971, 1972a, 1972b, 1973a, 1973b) explores this idea.

In a recent article in this journal, Dedrick (1998a) has argued that “some of the common objections to the works of Berlin and Kay and Rosch . . . are not significant” (p. 181).<sup>2</sup> Although he acknowledges that “the original research in *Basic Color Terms* (Berlin and Kay [1969] 1991) was flawed,” he argues that “the foundations of the tradition are . . . solid” (p. 181). What is solid is that “focal colors have a natural psychological salience” (p. 180). We disagree.

First, in the publication Dedrick (1998a, 181) criticizes (Saunders and van Brakel [1997] and van Brakel [1993]) the position that “colour naming is properly explained in strictly cultural terms” is not defended. Nor do these publications make a “contribution to a debate about what is natural and what is cultural in human colour categorization” (p. 182). Furthermore, they do not argue that “the significance of the colour-naming issue has got to do with ‘the nature and extent of relativism’ ” (p. 196). Second, Dedrick criticizes a number of brief passages in two subsections of Saunders and van Brakel (1997). Arguments he gives to support his criticisms are in fact forcefully addressed in other parts of that publication.<sup>3</sup>

Moreover, all of Dedrick’s substantial criticisms are based on simplifications and misreadings. For example, Saunders and van Brakel (1997, 169) reads, “Finally, although Berlin and/or Kay published var-

ious emendations to their theory, in particular to introduce more possible evolutionary sequences (Berlin and Berlin 1975; Kay 1975; Kay, Berlin, and Merrifield 1991), they have never addressed issues raised by their critics." After quoting this passage, Dedrick says this "blanket assertion . . . is vague, misleading, and in some cases simply false." In support, Dedrick lists references to work that does support Berlin and Kay. But the fact that others address some of the criticisms is irrelevant to the question whether Berlin and Kay themselves have addressed the issues. Dedrick then quotes a passage from Berlin and Kay ([1969] 1991) where the possibility of two BCTs for red in Hungarian and two BCTs for blue in Russian is mentioned. He seems to construe this as a rebuttal *avant la lettre*. However, a passage in the original work cannot falsify the statement that they have not addressed issues subsequently raised by their critics. Dedrick offers no further examples of Berlin and Kay's response.

Dedrick mentions that Berlin and Kay adjusted their theory in the light of "communications received from field linguists," and he gives one example (Dale Kinkade and Mary Haas [not "Mary Hass"]<sup>4</sup>). What we meant in the passage he calls a "blanket assertion" is that Berlin and Kay never responded to critics like Simpson (1991), Durbin (1972), Hickerson (1971), Ratner (1989), Shweder and Bourne (1984), McNeill (1972), Kuschel and Monberg (1974)—all of whom he refers to (Dedrick 1998a, 184, 186, 187).<sup>5</sup> When MacLaury (1992) published a target article in *Current Anthropology*, containing many criticisms of Berlin and Kay and presenting a new model based on his Meso-American Color Survey and the World Color Survey (WCS), Berlin and Kay chose not to respond.<sup>6</sup>

Of Berlin and Kay's comment on Russian blue and Hungarian red, Dedrick (1998a, 183) says, "The fact that Saunders and van Brakel (1997) missed this response is perplexing given that Hardin (1993) quoted the preceding passage from Berlin and Kay [about Russian blue and Hungarian red] in a recent discussion of van Brakel's (1993) own 'The Plasticity of Categories.'" What is perplexing is that

- Dedrick didn't notice that since 1988, Saunders and van Brakel have referred in print to the Russian blue and/or Hungarian red example five times;<sup>7</sup>
- Dedrick missed van Brakel's (1994b) response to Hardin (1993) in which 300 odd languages are listed that present problems for Berlin and Kay ([1969] 1991);<sup>8</sup> and
- Dedrick misses the point of "a whole industry developed to determine how many BCTs Russian has for blue (1 or 2?) and purple (0, 1, or 5?)"

(Saunders and van Brakel 1997, 169). He assumes that this is adduced as evidence for the fact “that there may be more than 11 basic colour terms” (1998a, 183). But the full passage reads, “Moreover the defining criteria for BCTs were extremely plastic. [Followed by 11 references.] For example, a whole industry developed to determine how many BCTs Russian has for blue (1 or 2?) and purple (0, 1, or 5?).” To spell it out, the sentence Dedrick takes to be about there being perhaps more than 11 BCTs is about the plasticity of defining criteria for BCTs.

### ELIMINATION OF EVOLUTION

Dedrick (1998a) continues by commenting on the phrase “more possible evolutionary sequences.” He says, “it is important to realise that Berlin and Kay have not added stages to their account” (p. 183). However, no one has suggested that Berlin and Kay added more stages. What is meant is that in the 1969 version there are two alternative sequences: at stage III (with BCTs for WHITE, BLACK, and RED) either YELLOW or GREEN emerges. In Kay, Berlin, and Merrifield (1991) and Kay et al. (1997), the number of possible sequences rises because of an increase in the number of composite categories. These now include red-white-yellow (stage I, II), green-blue-black (stage I, IIIb), red-yellow (stage II, IIIa), green-blue (stage IIIa, IIIc, IVa), yellow-green-blue (stage IIIc), white-yellow (stage IIIc), yellow-green (stage IIIe, IVc), blue-black (stage IIIe, IVb).<sup>9</sup>

Dedrick (1998a, 201, n. 1) asserts that he “will not consider Berlin and Kay’s evolutionary ordering,” because he has discussed it elsewhere. There he (1996, 513) writes, “one may be struck by how *disorderly* all this ‘evolution’ seems. . . . It could be that there is no strict developmental sequence that must be observed by all languages.” But whatever sort of interesting “regularities” would remain if the evolutionary ordering were dropped is quite unclear. Dedrick (1996) provides a battery of arguments to show that Kay, Berlin, and Merrifield (1991) offer no serious support for their belief that (linguistically defined) composite categories have been constructed from “primary” or “basic” or “fundamental” psychobiological building blocks. Referring to Witkowski and Brown (1977, 54) who made the same point, he says, “there is no science at all—physiological or psychological—that suggests that composite colour categories are salient in any domain but the linguistic” (Dedrick 1996, 514; cf. 517).

Dedrick appears to reason as follows:

- all attempts of Berlin and Kay to explain the biological or neurophysiological basis of the composite categories found in the World Color Survey are hopelessly flawed, but
- regularities are found, therefore they need another explanation.

Offering no alternative explanation he fails to say what—if not Evolutionary Stages—the “composites” are. (They constitute Evolutionary Stages 1-4.) To separate BCTs from the thesis of evolutionary typology is to drive a car without wheels. Dedrick thinks he can do this by distinguishing the regularities proposed by Berlin and Kay from their explanation. But at no point does he make clear what these “explanation-free” regularities are. Without the concept of evolution, there would be no interesting data to explain. There would merely be data showing that Hering’s opponent *Urfarben* white/black, red/green, and yellow/blue are not universally expressed in the languages of the world.<sup>10</sup>

#### THE WCS

In section 3 of his article, Dedrick endorses virtually all criticisms leveled at Berlin and Kay ([1969] 1991). We agree with him that “Hickerson’s (1971) review is comprehensive in this regard” (Dedrick 1998a, 184). But he dismisses it because “this low methodological level” is of no great importance to the debate between the universalist and relativist. Instead he says, “Let us settle for the claim that there is a rather large body of research that has been conducted in the past 25 years that supports, or is believed to support in some measure, the claims of Berlin and Kay” (p. 185). We are at a loss to understand his own methodology here. He says there are now “hundreds of informant interviews and dozens of detailed studies of particular languages and their colour vocabularies” (p. 185), referring to the WCS. But nothing follows from this: the number of cases that pose problems for the Berlin and Kay paradigm only increases and its philosophical and methodological weaknesses become more apparent.<sup>11</sup>

If details of the WCS are examined, one discovers that the project, “nearing completion” (Kay and Berlin 1997), actually began in 1976-77.<sup>12</sup> According to MacLaury (1986, 4), it was also “nearing completion during 1979.” It appeared as an unofficial report in 1985 but subsequently was “being revised” (MacLaury 1987, 120). Some results were then presented at the 1989 Meeting of the American

Anthropological Association and were published (Kay, Berlin, and Merrifield 1991). Copies of the WCS data sheets were put on sale by the Summer Institute of Linguistics (Dallas, Texas). They were later recalled on the grounds they were “only preliminary notes of the WCS project” that “were distributed prematurely and in error” and “which should not become part of any published record.”<sup>13</sup> If all previous work was based on “preliminary and partial analyses” and “the data contained errors, including coding errors which have now been corrected” (Kay and Berlin 1997), then whatever cleaned up and streamlined version is to be published can only be self-serving.<sup>14</sup>

If one looks at the methodology of the WCS,<sup>15</sup> the only real difference with Berlin and Kay ([1969] 1991) is more data. In the “bare data set” of the WCS (Berlin, Kay, and Merrifield 1991), for each of the 110 languages there are roughly 10 to 16 data sheets. These contain processed results from standardized experiments—a compressed version of the 1969 procedures. That is, color words are elicited by presenting subjects with the “universe” of individual, highly saturated Munsell color chips. In the data made public, no description is given of the procedures followed.<sup>16</sup> In general, it seems that most speakers could use any word they liked to name the chips, although some sheets contain information to the effect that certain words were disallowed.<sup>17</sup> In the original Berlin and Kay procedure, after the BCTs had been established (by the experimenter), speakers were asked to map the areas on the Munsell color chart to which their BCTs referred. But in the WCS, such mappings do not seem to have been carried out.

At all stages of processing, “noise” is eliminated so that the theory can handle uncontaminated data. For example, in 20 of the 110 languages of the WCS, individual speakers use the same word in response to white and black chips. So where a normal speaker of Waorani (Ecuador) uses a cognate of a term *waimo* for white, speaker 11 uses it for both black and white. Similarly, in Jicaque (Honduras), 8 of 10 speakers use *pje* to name white, whereas 2 speakers use it for both black and white. For Berik (Indonesia), *sinsini* WHITE is used by three speakers “as the name for dark or even black chips,” while another Berik speaker focused BLACK in both black and white. The WCS procedure suggests that such speakers are mad or have eye defects, because they regard black and white as “the same.” It isn’t possible to say that the term used anomalously doesn’t mean “white” (or “black” as the case may be), because it (presumably) has passed all “operational criteria” for a BCT.



Although all 110 sets of data contain a column titled “field gloss” (for so-called indigenous color words), this column is often empty. When field glosses are given, some are straightforward color words, but a variety of other glosses occur that at a later stage of processing are graded as “basic” or “nonbasic” color words. For example, in Kemtuik (Indonesia), *dark weather* becomes BLACK and basic, *kind of tree* becomes “purple” and nonbasic; in Kuku-Yalinji (Australia), field gloss *unripe* becomes GREEN, although labeled GRUE in the stage assignments; in Mantjiltjara (Australia), field gloss *blood-blood* becomes RED, and *earth* becomes BROWN as a WCS gloss, although the latter is also referred to as “pink” (in the stage assignment of speaker 7). The criteria for assessment seem completely ad hoc.

Such data (and the methodological tricks that create regularities) open a rift between the epistemological desiderata (the claim that “green” is a color) and the ontological presuppositions (a world independent of the form of all experience). The methodology of the WCS is an approach in which informants are processed through color-naming machinery to provide data to be distributed over a priori pigeon holes (Saunders 1999, 2000); anything to do with the real language, culture, settings, procedures, or people involved is removed (Lucy 1997). To read WCS data sheets is to be confronted with a well-defined system of operations floating freely outside any human activity. It is a system of operations detached both from its own origins and means of production, as well as from the origin and “means of production” of its “input.”

There is of course a great deal of order in the WCS, partly apparent and partly real.<sup>18</sup> Some of the order is created by the equipment and method used (the Munsell system),<sup>19</sup> which limits possible responses in such a way that nothing is measured but degree of compatibility between subject and experimenter. Order is also created by data processing (as illustrated above).<sup>20</sup> The result is the manufacture out of a medium and method of truncated metaphysical monsters. Nobody disputes that the majority of languages contain a word that in some contexts can be translated as “red” and that this has something to do with what all ecological systems “humans-environment” have in common. This and the observation of Berlin and Kay’s blue-and-green and red-and-yellow composites (or the later added yellow-or-green) were already noted in the 19th century by Magnus (1880), Allen (1879), and others. What then is the nature of the regularities Dedrick is after?

## THE COLOR DOMAIN

Discussing Berlin and Kay's definition of BCTs, Dedrick (1998a, 187) writes,

Kuschel and Monberg [1974] were unhappy with [the] assertion by their informants [viz., "that all the contextualised colour terms can be categorized under the basic colour term headings"], suggesting that it "may be the result of our proddings into a cultural category which is not experienced as a separate entity by the Bellonese themselves" (p. 241). We should note, however, that every language that has been studied appears to have some basic terms. Some languages have many terms and some have few, but all appear to have a scheme for describing colour in a disembodied way.

The structure of his argument seems to be as follows:

- Fundamental premise: every language studied appears to have some BCTs.
- Observation: Kuschel and Monberg (1974) express doubts that this is true for Bellona.
- Conclusion: Kuschel and Monberg (1974) are wrong; every language must have some BCTs.

The possibility that there might be a language without BCTs is not imaginable: there is a "universal biolinguistic disposition for the development of certain 'basic' colour words" (Dedrick 1998a, 179). Let's look at the Bellona case in more detail. (The Bellonese live on an island in the Pacific, not, as Dedrick [p. 184] says, on the Pacific Northwest Coast of Canada.)

When Kuschel and Monberg (1974) arrived to do fieldwork in 1971-72, Bellona seemed to have seven BCTs: *ungi* (black), *unga* (red), *susungu* (white), *sesenga* (violet, brown), *hengohengo* (yellow), *'usi'usi* (blue), and *sinusinu* (green). However, the language was rapidly changing due to contacts with the Western world. Kuschel and Monberg's aim therefore was to draw on memories prior to the first European settlements in 1938. After elaborate tests with the Munsell chart and extensive conversations, their main conclusions were as follows:

1. Although the speakers of Bellona have one of the most elaborate vocabularies of the Polynesian languages, they do not have a word for *color* and seem uninterested in it as a separate domain.<sup>21</sup> Nevertheless,

they employ a wide range of words in situations where Westerners might use color words; although rather like color words, none qualifies as a BCT.<sup>22</sup>

2. Munsell color chips do not provide tools sufficiently flexible to reveal either the actual use of color terms in Bellona or the relation between "pure color" (as exemplified by the Munsell system) and other aspects of appearance and/or its evaluation.<sup>23</sup>
3. Bellona color terms can be processed as a stage II language. Alternatively, it might be "considered to have an extremely sophisticated system of colour notation, with innumerable 'colour words,' way beyond the Western system, and thus much more sophisticated" (Kuschel and Monberg 1974, 241).

Viewing Kuschel and Monberg's data in terms of Berlin and Kay's evolutionary sequence, a more straightforward conclusion might be that the Bellonese are at Evolutionary Stage 0: they have no BCTs. However, Kuschel and Monberg reluctantly attributed Evolutionary Stage II to them. This was because when pressed, informants distributed all color chips over three "mothers of color," although these words could not always be used in cases which, by Western standards, were salient cases of white, black, or red.<sup>24</sup> The conclusion suggests itself that the three mothers of color (BCTs) were ad hoc coinings in the experimental situation. First, the Bellona were goaded/encouraged/forced to use a limited number of words/categories to name continuous patches on the Munsell color grid (in other words, they were made to engage in a forced-choice experiment). Second, globalizing influences (schooling, missionaries, linguists, anthropologists) had already provoked a reflexive stance that made them think they ought to have mothers of color.

## ROSCH

Saunders and van Brakel (1997) list six criticisms of the work of Rosch. Dedrick (1998a, 193) discusses a small part of one of them. The part he quotes reads,

[Rosch's] experiments *did* show that the Dani remembered focal colours better than non-focal ones, as did Americans. However when asked to point out a focal colour shown 30 seconds before in an array of 160 colours, Dani people were mistaken 75% of the time, Americans 34%. If humankind has a biological sensitivity to focals, it is difficult to understand how this level of error, or the difference between the Dani and Americans, can be explained. (P. 170)

After quoting this passage, he says, "Saunders and van Brakel failed to inform their readers of a crucial methodological presupposition of Rosch's work with the Dani" (Dedrick 1998a, 193). But in the other five criticisms, Rosch's methodological presuppositions are fully addressed. Moreover, what Dedrick quotes is not about the Dani scoring worse than American students in a typical psychologist's lab experiment. The point is merely rhetorical: if there are natural psychological saliences, why can't the Dani, let alone bright American students, keep them in their heads for 30 seconds? So, the crucial figure is that the Americans scored only 66%.

Nevertheless, Dedrick (1998a, 195) enters into an elaborate thought experiment to prove that "there is a natural psychological salience that explains the similarity in performance." But this thought experiment begs the question by assuming that "the universally salient stimulus of unique green" means the same as "what-we-around-here-call-a-good-green." More important, his conclusion amounts to saying that regardless of which group of normal humans one takes, all are predisposed to remember or re-recognize a particularly deep, lustrous, and succulent green when placed among some boring desaturated greens. It was in trying to preempt muddles like Dedrick's thought experiment that part of the fourth criticism of Rosch reads,<sup>25</sup>

What was confirmed in Rosch's experiments (with Dani people and others) was the primacy of *focal* colours defined by *saturation*. . . . Given a particular hue category (for the Dani: as defined by Rosch in terms of three colour chips), it would seem self-evident that the best example is the most saturated, because "most saturated" means "having most colour." (P. 182)

The main problem is that Dedrick's account of Rosch's work is familiar hagiography. We don't have space here to critically review all aspects of her work on color, but the following brief remarks may suggest caution.<sup>26</sup>

In 1968, Rosch first visited the Dani of Irian Jaya, where her husband K. G. Heider was doing fieldwork (Heider 1970, 175; Rosch 1988, 376). Though working in the Brown and Lenneberg (1954) tradition (of language and cognition studies), she had already proposed that color recognition and memory were not mediated by language (Rosch 1988, 376). The Dani were "ideal subjects" for experiments because they "were reported to have only two color terms": *mili* and *mola*.<sup>27</sup> Returning to America, she read Berlin and Kay ([1969] 1991) in

manuscript (Rosch 1988, 377). Brown and Lenneberg's (1954) "most codable" colors were clearly Berlin and Kay's "clusters of foci." She was convinced at that point that the structure of reality (i.e., the foci), not the structure of language, determined cognition.

On the model of Berlin and Kay ([1969] 1991), the translation of the two Dani color terms *mili* and *mola* should have been BLACK and WHITE (Berlin and Kay [1969] 1991, 46-47). But that is not what came out of Rosch's experiments. Rather, "the ethnographer must see simultaneously that a very dark 'warm' and a fairly light 'cold' colour are both called *mola*" (Rosch 1972b, 461-62).<sup>28</sup> Kay (1975, 258) later streamlined her data. He claimed, "such (two-term) systems contrast dark and cool hues on the one hand against light and warm hues on the other." This casual disdain for others' results characterizes the Berlin and Kay program then as now.

In her only anthropological publication, Rosch warned that "the ethnographer's intake of information is always a kind of sampling, however informally or unconsciously guided" (Rosch 1972b, 448). But when it comes to her own fieldwork, she can't live up to such wisdom. Describing her work (Rosch 1972a), she says, "in 23 diverse languages, drawn from seven of the major language families of the world, it was the same colors that were most codable" (Rosch 1974, 111). If one looks how this random drawing of languages was done, it turns out that the experiment was carried out with 14 bilinguals in the United States and 9 bilinguals in Indonesia. Moreover, of the 23 native languages of these bilinguals, at least 18 were at what Berlin and Kay called stage VII, that is, they had the same number of BCTs as English.<sup>29</sup> This reproduces the problem of Berlin and Kay's ([1969] 1991) original study in which 15 out of 20 languages were at stage VII. This is one example of why we find it difficult to put much weight on her results.

#### THE LINGUIST AS LANGUAGE MAKER

A plausible explanation for the ubiquity of common descriptive color meanings in 20th-century languages is the spread of similar historical and technological conditions. For example, artificial dyes, global color standards, and a quasi-scientific color language guarantee commensurability.<sup>30</sup> The impact of such conditions destroys local practices, either by slow penetration or by elimination. Linguists have been especially instrumental in this process.

As an example, let's look at the Pacific Northwest of Canada. Vancouver ([1798] 1984, 281, Vol. II) noted "the very lively and beautiful yellow" that (among other hues) bordered finely wrought garments of "pine bark." Hence, there was precontact knowledge and practice of yellow dyes.<sup>31</sup> But, few Northwest Coast languages have a separate word for yellow. In Kwakw'ala, for example, there is one word *thenxa* for what is called either green or yellow in English. The anomaly might be expected to disappear if speakers knew that yellow and green are two different unique colors, two of the four neurophysiologically built-in opponent hues. But, though most contemporary speakers of Kwakw'ala are bilingual and know perfectly well the difference in English between yellow and green, they stick to *thenxa* in Kwakw'ala. We are inclined to take such counterexamples seriously. One question it raises is, Why do the "innate cognitive categories" always coincide with quasi-scientific American English? In the Berlin/Kay/Rosch tradition, American English functions as the metaphysically neutral metalanguage. Rather than saying that American English has hit on the truth, we are inclined to say that the BCTs have been molded by scientific theory and technological practice.

Moreover, talking about the Kwakw'ala language is itself a linguistic invention, constructed out of an indefinite number of dialects and idiolects. As early transcription started at Fort Rupert (on the north end of Vancouver Island) and many people had moved there because of its economic significance,<sup>32</sup> the Kwakw'ala language was constructed out of the local dialect, primarily by the efforts of the anthropologist Franz Boas and George Hunt, who lived at Fort Rupert. By the first decade of the 20th century when Boas began his analysis of the "grammatical concepts" of Northwest Coast languages, in particular Kwakw'ala, the Kwakiutl (now called the Kwakwaka'wakw) had been influenced by contact with European conditions and a settler society for more than 125 years. For a language that he later described as deeply metaphorical and religious (Boas 1930, 1931), he struggled throughout the 1890s with the difficulties of identifying "a word" as a meaningful unit of translation (Boas 1891, 1892). Moreover, he was confronted by what seemed to be particles and predicates rather than nouns and verbs (Boas 1947). By 1905, however, he seems to have resolved these difficulties by using "roots" or "stems" that he took to be "core morphemes" (cf. Boas 1911). Using this method, Boas and Hunt ([1905] 1975) gave the definitive inscriptions of the basic Kwakw'ala color morphemes: *ts'olh*- "black," *mel*- "white," *flaqw*- "red," *thenx*- "green," *dzas*- "blue," and *moqw*- "yellowish." In 1990,

however, no Kwakw'ala speaker had heard of *moqw*<sup>33</sup> but all knew *lhenxa* and insisted there was no separate term for "yellow" (Saunders 1992)—although all knew "yellow" and "green" in English.

The earliest written source on Kwakw'ala (Dawson [1887] 1973) glossed *lhenxa* as "green, yellow." However, from 1892 to 1934, there are five publications (Boas 1892, 1931, [1934] 1969; Boas and Hunt [1905] 1975; Curtis [1916] 1975) that list *lhenxa* as "green." In the 1970s and 1980s, *lhenxa* was again glossed as "green, yellow" (Grubb 1977; Lincoln and Rath 1980). If for some reason the Kwakw'ala language would have disappeared as a spoken language by 1950, then Dawson ([1887] 1973) would have been forgotten and there would merely be confirmation of the universality of BCTs.

The processes surrounding the inscription of Kwakw'ala color words in the period of 1892 to 1934 epitomize the historical forces. Functional hypertrophied vision and a technologically and economically mediated mode of reasoning standardized perception. Once inscribed, the complex polymorphic interactions were reduced to a single "correct" mode of perception, itself the result of historical and technological processes (Saunders 1998, 1999, 2000).

This "rewriting of color" is not only found in Berlin and Kay ([1969] 1991)<sup>34</sup> and the WCS but also in a host of other psychological and linguistic works. For example, Uchikawa and Boynton (1987) and Boynton and Olson (1987), using OSA (Optical Society of America) instead of Munsell confirm Berlin and Kay's thesis of 11 BCTs for American and Japanese speakers—a result quoted approvingly by others (for example, Hardin 1993; Pokorny, Shevell, and Smith 1991). Historical language change is irrelevant to this paradigm. Yet, it is only since 1860 when synthetic dyes based on materials such as coal tar and petroleum were invented and imported by Japan that the language took on its basic color terms. Even today, another color naming system exists. As Stanlaw (1987) makes abundantly clear, there is a difference between traditional and modern Japanese.<sup>35</sup> High-tech experiments measuring the names of Munsell or OSA color patches merely solicit responses in modern/Western Japanese.

## CONCLUSION

By urging that BCTs be regarded as separate from their evolutionary moorings and that Rosch et al. and the WCS reconfirm the Berlin and Kay program, Dedrick does much to obfuscate efforts to rethink



it. Failing to provide evidence for regularities, he merely stipulates perceptual biological grounding. Were he to open that up, he would find another arena of contestation and controversy.

The color paradigm Dedrick supports commits a standard epistemological solecism in the human sciences. It places the model the scientist constructs in the brain of the subject and takes it to operate as if that construct were the main causal determinant of linguistic and categorical practice. The reality of the model is confounded with the model of reality and that in turn is presented as a higher form of reality.

### NOTES

1. In a language with three basic color terms (BCTs), RED agglomerates red, yellow, orange, pink, and purple. WHITE includes all light hues. BLACK includes blue and green. In a language with five BCTs, GREEN covers greens and blues until the sixth BCT, blue, emerges.

2. See also the appendix in Dedrick (1998b).

3. See, for example, his reference to the yellow-green category (Dedrick 1998a, 183), his appeal to simplicity (p. 189), his reference to the World Color Survey (WCS), and discussion of the Bellona data. These issues are discussed in sections 4.3 and 6.2 of the target article and in the "Author's Reply" of Saunders and van Brakel (1997).

4. Adjustments to the evolutionary sequence in Kay (1975), Kay and McDaniel (1978), and Kay, Berlin, and Merrifield (1991) were made partly in response to publications of Rosch, MacLaury, and others.

5. Kay and McDaniel (1978) refer to Kuschel and Monberg (1974) as part of "evidence [that] has been adduced supporting the hypothesised sequence of temporal development in basic color-term systems" (p. 615).

6. Kay, Berlin, and Merrifield (1991) acknowledge the earlier work of MacLaury, which they incorporate in their revised evolutionary sequence.

7. Saunders (1992, 28, n. 55), Saunders and van Brakel (1988, 366; 1995, 163, n. 6), van Brakel (1993, 114; 1994b, 773).

8. van Brakel (1994b) should be consulted in conjunction with van Brakel (1993). The 1994 article basically consists of three tables and 336 references. It was published in response to Hardin's (1993, 144) complaint that "van Brakel (1993) douses us with a plethora of assertions about various linguistic tidbits . . . with only the scantiest suggestion of sources and evidence."

9. The number of possible evolutionary sequences increases further because derived terms (*gray, brown, orange, purple, pink*) can emerge at any stage as wild-card colors:

"The temporal development of basic color-term systems should be seen, not as a single process, but as two partially independent processes: (i) the division of composite categories into the six fundamentals, and (ii) the combination of fundamental categories into derived categories." (Kay et al. 1997, 29)



10. This was already known when Hering proposed his theory of opponent processes (Allen 1879; Magnus 1880).

11. Saunders (1992, 1995, 1999, 2000), Saunders and van Brakel (1995, 1997), van Brakel (1993, 1994b).

12. National Science Foundation grant BNS 76-14153. See Saunders (1992, 117).

13. Letter from Summer Institute of Linguistics to buyers of WCS data sheets, July 1993.

14. Kay et al. (1997, 34) announced “a two volume monograph” on the WCS. The second volume will be a “presentation of the WCS data in a format that will make them readily available to all scholars.” In a note they add, “The format described here reflects our current thinking on the monograph. These decisions are subject to revision as the work proceeds.”

15. Most of the examples given below are taken from Saunders and van Brakel (1995).

16. There is a brief description of the field procedures in an early preliminary report presented at the AAA (American Anthropological Association) in 1985, but this report is marked “not for quotation.”

17. For example, for Guahibo (Colombia), “The investigator did not allow the speakers to use term 12 [field gloss: ‘type of bird’; WCS gloss: ‘pink’] and 14 [field gloss: ‘type of clay’; WCS gloss: ‘pastel’] until the 21st speaker.”

18. For about a century, color has been the standard example in discussions of cultural universals and linguistic relativism. Surprisingly, in innumerable sophisticated philosophical, psychological, and linguistic publications, the relativistic position is explained with reference to the fact that “other people” have the custom of “dividing the color spectrum differently.” But neither the universalist nor the relativist seems to notice that in disagreeing on whether other people divide the spectrum differently, both agree that color is a universal.

19. The Munsell system was built on the basis of Maxwell discs, Fechner’s just noticeable differences and Hering’s opponent processes, organized according to the three dimensions of hue-saturation-lightness. Soon it was reduced to a system of samples and stimuli, which were plotted onto the physical CIE (Congès International d’Éclairage) system to smooth out irregularities (Sivik 1997; Johnston 1996; Simpson 1991). Dominant wavelength, intensity, and excitation purity—the physical analogons of hue, lightness, and saturation—were calculated for nominal Munsell notations to create conversion tables, and all Munsell codes were tied to CIE tristimulus values (Wyszecki and Stiles [1967] 1982).

20. Compare the hoax to fit the Fore people into Ekman’s theory of universal facial expressions for basic emotions (van Brakel 1994a, 189-91).

21. For example, color words were rarely used to teach children properties of plants, fish, and so on; few color words appear in myths; the Bellona use no color symbolism with respect to sex, mourning, and so on; both men and women were far more interested in patterns than colors.

22. Although in precontact Bellona there is a sense in which *unga* means red, this has to be heavily annotated. There are numerous red objects that cannot be called *unga*. For example, the red skin surrounding the kernel of the tangie fruit is not *unga* but is *koka*—*koka* is primarily used to describe the process of dyeing wooden objects with a reddish color derived from the skin or roots of the *Morinda citrifolia*. There are different words for the redness of feathers (*kunga*), the brownness of a certain type of coconut (*keka*), the redness of vulgar (and some other) objects (*mea*), whereas red-colored skin

and fur is usually *segħa*. In all these cases of red and many more, *unga* cannot be used. A reasonable gloss of *unga* is "red and some other bright colours, but not (i) for most skins, feathers, furs, etc., (ii) for changing colours, (iii) if the colour is valued positively."

23. Often, Bellona words did not refer to color as the property of an object but to a process of change, such words being frequently evaluative. In general, Bellona color words are used for a range of properties, objects, situations, or events (of degrees of abstractness or particularity), which to Westerners appear disparate and unconnected. For example, one of the many words for blackness or darkness, *lalangi*, could be used of a dark night, black tattoos, flying foxes, and Melanesians from the West Solomons but not of hair, whales, or fish.

24. See note 21.

25. Compare Dedrick's (1998a, 182) observation that "Collier . . . demonstrated that saturation is indeed a key aspect of focality."

26. For more details, see Saunders (1992, chap. 6).

27. Although linguists and ethnographers have pointed to complexities of the Dani language (van der Stap 1966; Heider 1979) suggesting that *mili* and *mola* are evaluatory words (Heider 1970, 175f), Rosch herself merely noted that the "Dani Ss tended to 'chant' the two names at a constant rate" (Rosch 1972a, 16).

28. In several other publications, Rosch reports that the Dani "divide the color space on the basis of brightness rather than hue" (1972a, 15; 1973a, 331; 1973b, 115; 1974, 115). Compare also Rosch and Olivier (1972, 340, 343): "brightness based," "light/warm"; Rosch (1972b, 451): "not based purely on brightness"; Rosch (1973a, 332): "the two basic Dani color terms 'mili' (roughly 'dark') and 'mola' (roughly 'light')." Kay, Berlin, and Merrifield (1991) go further, saying, "stage I, with two composite terms—white/red/yellow and green/blue/black" is the one case "extensively documented by E. Rosch" (Rosch 1972a, 1972b).

29. Of the languages Rosch lists, only Yoruba and Batak are not stage VII in the sense of Berlin and Kay ([1969] 1991). Of three of the languages mentioned by Rosch, we have no information: Birmese, Hausa, and Sundanese. Sundanese, like Batak, Bahasa Indonesia, and Javanese is an Indonesian language. The data for Batak in Berlin and Kay ([1969] 1991) are derived from Bartlett (1929), who says that their word for yellow, orange, and brown is *hoenik* or *koening*, which is derived from Dutch *honing* (honey). It is plausible that by the time Rosch did her experiments, Batak was, by Berlin and Kay standards, a stage VII language. This might apply to Sundanese too.

30. Consider, for example, Khmer (Cambodia) *sukula* and Gujarati (India) *shoklati*; that is, brown, which somehow got there from Spanish (perhaps via Tagalog in the Philippines), whereas Spanish got the word from Nahuatl (Mexico) *chokolatl* (food made from cacao seeds). In return, Nahuatl borrowed *kafentik* (for brown) from Spanish *café* (coffee), where it came via Turkish and Arabic from (probably) Ethiopia.

31. For more details, see Saunders (1992).

32. Lucrative coal mining was the main economic motor.

33. According to Lincoln and Rath (1980), *muqwa* is used in the northern areas where it is one of three root morphemes for "white." Perhaps *mukwa-* belongs to a series of "shining," "white," "light," "bright," "glitter," and "sparkle" words. Boas (1892) gives *mu'k'ola* "moon"; Boas and Hunt ([1905] 1975) give *moxwp'eq* "pine."

34. As Hickerson (1971) pointed out, implicit rules were used to assign meanings to BCTs in a particular language on the basis of written entries. These rules guarantee the theory will be confirmed. Given a list of BCTs in a language, first the BCTs WHITE and BLACK are selected, then RED. Where there is no appropriate gloss, a correction is

made, as in Poto: *eyeyengo* "yellow" → RED. The next stage must either be GREEN or YELLOW; hence, for Daza: *zede* "jaune, bleu, vert" → GREEN and Arunta *tierra* "yellow, green, blue" → YELLOW. Similarly, in Mazatec *sase*, glossed as "blue, blue-greens, blue-violets" is assigned GREEN, because GREEN necessarily appears before BLUE. It has been suggested (Hardin 1993) that this procedure is permissible because the capitalized words are not glosses but names of prelinguistic basic color categories. But that is of course the point at issue and thus begs the question. What is going on is that the content of the BCTs is taken as given. Data have no relevance as possible supporters or falsifiers; they are simply fitted to the a priori model.

35. Compare also Hickerson (1971, 262) who found Berlin and Kay's reconstruction of Japanese superficial, and Wierzbicka (1990) and Stanlaw (1987) who point out that Berlin and Kay's BCTs for blue and green, *ao* and *midori*, are not mutually exclusive but seem to overlap completely on a hue scale, although they cannot be used as synonyms in all contexts. Moreover, apart from these BCTs for blue and green, modern Japanese seems to have the BCTs *kimidori* "chartreuse"; *mizu* "light blue, aquamarine, water color"; and *kon* "dark blue" (Iijima, Wenning, Zollinger 1982; Johnson and Tomiie 1985; Stanlaw 1987; Uchikawa and Boynton 1987).

## REFERENCES

- Allen, G. 1879. *The colour sense: Its origin and development*. London: Trubner.
- Bartlett, H. H. 1929. Color nomenclature in Batak and Malay. *Michigan Academy of Science, Arts and Letters* 10:1-52.
- Berlin, B., and E. Berlin 1975. Aguaruna color terms. *American Ethnologist* 2:61-87.
- Berlin, B., and P. Kay. [1969] 1991. *Universality and evolution*. Berkeley: University of California Press.
- Berlin, B., P. Kay, and W. R. Merrifield. 1991. *The World Color Survey*. Photocopied data available from the Summer Institute of Linguistics, Dallas, Texas.
- Boas, F. 1891. Second general report on the Indians of British Columbia. In *Report of the British Association for the Advancement of Science 1890*, 562-715. London: Murray.
- . 1892. Vocabulary of the Kwakiutl language. *Proceedings of the American Philosophical Society* (Philadelphia: Mac Calla) 30:34-82.
- . 1911. Kwakiutl. In *Handbook of American Indian Languages*, 425-557. Smithsonian Institution Bureau of American Ethnology Bulletin 40, Part I. Oosterhout: Anthropological Publications.
- . 1930. *Religion of the Kwakiutl*. Columbia University Contributions to Anthropology 10, Part 1. New York: Columbia University Press.
- . 1931. Notes on the Kwakiutl vocabulary. *International Journal of American Linguistics* 6:163-78.
- . [1934] 1969. *Geographical names of the Kwakiutl Indians*. New York: AMS Press.
- . 1947. Kwakiutl grammar with a glossary of the suffixes. *Transactions of the American Philosophical Society* 37 (3): 203-377.
- Boas, F., and G. Hunt. [1905] 1975. *Kwakiutl texts*. (Jesup North Pacific Expedition, Vol. 3; Memoir of the American Museum of Natural History, New York, Vol. 3; Leiden: Brill, New York: Stechert). New York: AMS Press.
- Boynton, R. M., and C. X. Olson. 1987. Locating basic colors in the OSA space. *Color Research and Application* 12:94-105.

- Brown, R., and E. Lenneberg. 1954. A study in language and cognition. *Journal of Abnormal and Social Psychology* 49:454-62.
- Curtis, E. S. [1916] 1975. *The North American Indian*. Vol. 11. New York: Johnson Reprint.
- Dawson, G. M. [1887] 1973. *Notes and observations on the Qwakiool people of the northern part of Vancouver Island and adjacent coasts, made during the summer of 1885*. Fairfield, WA: Ye Galleon Press.
- Dedrick, D. 1996. Colour language universality and evolution: On the explanation for basic colour terms. *Philosophical Psychology* 9:497-524.
- . 1998a. On the foundations of the universalist tradition in colour naming research (and their supposed refutation). *Philosophy of the Social Sciences* 28:179-204.
- . 1998b. *Naming the rainbow: Colour language, colour science, and culture*. Dordrecht: Kluwer.
- Durbin, M. 1972. Basic terms—Off-color? *Semiotica* 6:257-78.
- Grubb, D. McC. 1977. *A practical writing system and short dictionary of Kwakw'ala (Kwakiutl)*. Ethnology Service Paper No. 34. Ottawa: National Museums of Canada.
- Hardin, C. L. 1993. van Brakel and the not-so-naked emperor. *British Journal for the Philosophy of Science* 44:137-50.
- Heider, K. G. 1970. *The Dugum Dani*. Chicago: Aldine.
- . 1979. *Grand Valley Dani: Peaceful warriors*. New York: Holt, Rinehart and Winston.
- Hickerson, N. P. 1971. Review of Berlin & Kay (1969). *International Journal of American Linguistics* 37:257-70.
- Iijima, T., W. Wenning, and H. Zollinger. 1982. Cultural factors of color naming in Japanese: Naming tests with Japanese children in Japan and Europe. *Anthropological Linguistics* 24:245-62.
- Johnson, E. G., and T. Tomiie. 1985. The development of colour-naming in four- to seven-year old children: A cross-cultural study. *Psychologia: An International Journal of Psychology in the Orient* 28:216-27.
- Johnston, S. F. 1996. The construction of colorimetry by committee. *Science in Context* 9 (4): 387-420.
- Kay, P. 1975. Synchronic variability and diachronic change in basic color terms. *Language in Society* 4:257-70.
- Kay, P., and B. Berlin. 1997. Science is not imperialism: There are nontrivial constraints on color naming. *Behavioural and Brain Science* 20:196-201.
- Kay, P., B. Berlin, L. Maffi, and W. Merrifield. 1997. Color naming across languages. In *Color categories in thought and language*, edited by C. L. Hardin and L. Maffi, 21-58. New York: Cambridge University Press.
- Kay, P., B. Berlin, and W. R. Merrifield. 1991. Biocultural implications of color naming. *Linguistic Anthropology* 1:12-25.
- Kay, P., and C. K. McDaniel. 1978. The linguistic significance of the meanings of basic color terms. *Language* 54:610-46.
- Kuschel, R., and T. Monberg. 1974. "We don't talk much about colour here": A study of colour semantics on Bellona Island. *Man* 9:213-42.
- Lincoln, N. J., and J. C. Rath. 1980. *North Wakashan comparative root list*. Canadian Ethnology Service Paper No. 68. Ottawa: National Museums of Canada.
- Lucy, J. A. 1997. The linguistics of "color." In *Color categories in thought and language*, edited by C. L. Hardin and L. Maffi, 320-46. New York: Cambridge University Press.
- MacLaury, R. E. 1986. *Color in Mesoamerica*, Vol. 1: A theory of composite categorization. Ph.D. diss., University of California, Berkeley.

- . 1987. Color-category evolution and Shuswap yellow-with-green. *American Anthropologist* 89:107-24.
- . 1992. From brightness to hue: An explanatory model of color-category evolution. *Current Anthropology* 33:137-86.
- Magnus, H. 1880. *Untersuchungen über den Farbensinn der Naturvölker*. Jena: Fischer.
- McNeill, N. B. 1972. Colour and Colour Terminology. *Journal of Linguistics* 8: 21-23.
- Pokorny, J., S. K. Shevell, and V. C. Smith. 1991. Colour appearance and colour constancy. In *The perception of colour*, edited by P. Gouras, Vol. 6, 43-61. London: Macmillan.
- Ratner, C. 1989. A sociohistorical critique of naturalistic theories of color perception. *Journal of Mind & Behavior* 10:361-72.
- Rosch, E. 1971. "Focal" color areas and the development of names. *Development Psychology* 4:447-55.
- . 1972a. Universals in color naming and memory. *Journal of Experimental Psychology* 93:10-20.
- . 1972b. Probabilities, sampling, and ethnographic method: The case of Dani colour names. *Man* 7:448-66.
- . 1973a. Natural categories. *Cognitive Psychology* 4:328-50.
- . 1973b. On the internal structures of perceptual and semantic categories. In *Cognitive development and the acquisition of language*, edited by I. E. Moore, 111-44. New York: Academic Press.
- . 1974. Linguistic relativity. In *Human communication: Theoretical explorations*, edited by A. Silverstein, 95-121. New York: Halsted.
- . 1988. Coherences and categorisation: A historical view. In *Development of language and language resources*, edited by F. Kessel, 373-92. Hillsdale, NJ: Lawrence Erlbaum.
- Rosch, E., and D. C. Olivier. 1972. The structure of the color space in naming and memory for two languages. *Cognitive Psychology* 3:337-54.
- Saunders, B.A.C. 1992. *The invention of basic colour terms*. Utrecht: ISOR.
- . 1995. Disinterring basic color terms: A study in the mystique of cognitivism. *History of the Human Sciences* 8 (4): 19-38.
- . 1998. What is colour? *British Journal of Psychology* 89:697-704.
- . 1999. The spectre of colour: A sociobiological paradigm. *Science as Culture* 8:473-96.
- . 2000. Revisiting basic color terms. *Journal of the Royal Anthropological Institute* 6:81-98.
- Saunders, B.A.C., and J. van Brakel. 1988. Re-evaluating basic colour terms. *Cultural Dynamics* 1:359-78.
- . 1995. Translating the world color survey. In *Post-modernism and anthropology*, edited by K. Geuijen, D. Raven, and J. de Wolf, 161-78. Assen: van Gorcum.
- . 1997. Are there non-trivial constraints on colour categorisation? *Behavioral and Brain Sciences* 20:167-232.
- Shweder, R. A., and E. J. Bourne. 1984. Does the concept of the person vary cross-culturally? In *Culture theory: Essays on mind, self, and emotion*, edited by R. A. Shweder and R. A. LeVine. Cambridge: Cambridge University Press.
- Simpson, C. 1991. Colour perception: Cross-cultural linguistic translation and relativism. *Journal for the Theory of Social Behavior* 21:409-30.
- Sivik, L. 1997. Color systems for cognitive research. In *Color categories in thought and language*, edited by C. L. Hardin and L. Maffi. New York: Cambridge University Press.

- Stanlaw, J. M. 1987. Color, culture, and contact: English loanwords and problems of color nomenclature in modern Japanese. Ph.D. thesis, University of Illinois at Urbana-Champaign.
- Uchikawa, K., and R. M. Boynton. 1987. Categorical color perception of Japanese observers: Comparison with that of Americans. *Vision Research* 27:1825-33.
- van Brakel, J. 1993. The plasticity of categories: The case of colour. *British Journal for the Philosophy of Science* 44:103-35.
- . 1994a. Emotions as the fabric of forms of life: A cross-cultural perspective. In *Social perspectives on emotion*, edited by W. M. Wentworth and J. Ryan, Vol. II, 179-237. Greenwich, CT: JAI.
- . 1994b. The ignis fatuus of semantic universalia: The case of colour. *British Journal for the Philosophy of Science* 45:770-83.
- Vancouver, G. [1798] 1984. *A narrative or journal of a voyage of discovery to the north Pacific ocean and round the world. . . . 1790-1795 in the "discovery" sloop of war, and armed tender "Chatham" under the command of Captain George Vancouver*. 3 Vols. London: Robinson. (Quoted from 1984 reprint, W. K. Lamb, ed., Vol. 1, London: Hakluyt Society)
- van der Stap, P.A.H. 1966. *Outline of Dani morphology*. The Hague, the Netherlands: Nijhoff.
- Wierzbicka, A. 1990. The meaning of color terms: Semantics, culture and cognition. *Cognitive Linguistics* 1:99-150.
- Witkowski, S. R., and C. H. Brown. 1977. An explanation of color nomenclature universals. *American Anthropologist* 79:50-57.
- Wyszecki, G., and W. S. Stiles. [1967] 1982. *Color science*. 2nd ed. New York: Wiley.

## Review Essay

---

### Freeman on Mead Again

I. C. JARVIE  
York University, Toronto

Derek Freeman, *The Fateful Hoaxing of Margaret Mead. A Historical Analysis of Her Samoan Research*. Westview, Boulder, CO, 1999. Pp. xi+279. \$24.00.

In November 1983, the late Derek Freeman was anathematized by resolution of the 82nd Annual Meeting of the American Anthropological Association (AAA) for a book that was, in the words of the motion, "poorly written, unscientific, irresponsible and misleading" (208-9). Needless to say, Freeman was not present and was given no official forum in which to answer nor was there any process of appeal. He was the victim, in the strictest sense, of a kangaroo court. The book in question was *Margaret Mead and Samoa: The Making and Unmaking of an Anthropological Myth* (1983), published by no less than Harvard University Press. In it, Freeman argued that Margaret Mead's description of Samoan adolescent sexual mores in *Coming of Age in Samoa* (1928) was fundamentally flawed. Her picture of a period of free love under the palm trees for the unmarried Samoan adolescent was hard to reconcile with compelling direct evidence that Samoan society strongly emphasized premarital chastity, which was tested by a male relative. It also conflicted with the circumstantial evidence of the puritan Christianity to which Samoans adhered. So convinced was Mead that there were no adolescent sexual problems on Samoa, and that this resulted in enhanced sexual health, that in other publications she declared "the idea of forceful rape or of any sexual act to which both parties do not give themselves freely is completely foreign to the Samoan mind" (p. 187). Freeman's bloodhound instinct led him to examine court records, where he found evidence of an incidence of



rape twice that of mainland United States (in Western Samoa). In the book, he reports Mead recording two incidents of rape in her loose-leaf folder.

Freeman had put his doubts about Mead's 1928 findings to her directly and she had been nonplussed, wondering if perhaps what she had found was confined to that time and place, only to be different elsewhere and later. (Although Mead seems in 1925 to have been unaware of the large differences within the Polynesian "cultural area" between East and West.) Freeman, however, was not convinced by this suggestion. He too had done fieldwork in Samoa, was considerably more fluent in the language than was Mead, and like her, had been honored with a ceremonial rank. His expertise was as solid as was her good faith.

As Freeman admits in the present volume, he originally thought the issue was a purely scientific dispute between two scholars. Certainly, their own relations seem to bear that out. Freeman reports hesitating to publish and endeavoring to engage in further discussion with Mead, abortive only because she fell fatally ill. The result was that *Margaret Mead and Samoa* was received by an American anthropological community that revered Margaret Mead as standard bearer for the ideas of her teachers Franz Boas and Ruth Benedict, and that included some Samoa specialists for whom her work was exemplary.

Freeman the scientist was in for a surprise. His scientific claim, his competence, his underlying motives, and his integrity were all subject to attack. Mead's partisans were a good deal more *ad hominem* than she was herself. As one reads about the resistance to Freeman's claims and the attempts at rebuttal, it becomes clear that he had stumbled on a scientific dispute that involved identity and associated emotions. Freeman viewed Mead's error as important since it functioned as a crucial test of the underlying Boasian view that the form adolescence takes is a cultural particular rather than a universal (or biological) developmental phase. Boas had himself selected for Mead the problem of adolescence in Samoa as a crucial test of his view that culture was almost the whole story. Samoa as presented by Margaret Mead was thus what Bacon called an instance of the fingerpost: it had decisively pointed American anthropology down the path of culture and away from the path of biology.

A rational reception of Freeman's 1983 book would have required of American anthropologists that they call into question not just the truth of the researches of Margaret Mead and the ideas of her teachers but their entire cultural and nurturist identity. This they utterly re-



fused to do—as the condemnatory resolution of the 1983 annual meeting of the AAA shows. Just as the annual meeting had been used 15 years before to issue pronouncements on the rights and wrongs of the Vietnam war, it was in 1983 encouraged to settle, by a show of hands, intellectual issues that were matters of evidence and reasoning, not majority opinion. Although not offered a hearing, Freeman was quite able to defend himself in print, did so vigorously, and found allies among anthropologists, though the latter were mostly of antipodean and British allegiance rather than North American. If his carefully argued responses did not shame his denouncers, they should have.

So much for background. The book under review could be looked at like this. If there is any rational core to all the disagreement over Mead, then perhaps the issues in question deserve a second look. If the issues are at all responsive to evidence, then perhaps more of it will suffice to achieve rational closure. Revisiting the matter and marshalling further evidence carries the risk, of course, that Freeman might have to concede error.

*The Fateful Hoaxing of Margaret Mead* pursues general and particular projects of reevaluation. The general project is to bring together material that allows Freeman to reconstruct Mead's sojourn on Samoa almost day by day. He is also able to pinpoint just exactly what she was supposed to be doing and what she sometimes did instead. She was supposed to be concentrating on Samoan adolescence as a crucial test of Boasian culturalism—it was for this that Boas had secured her research grant. What she did instead was to give that project relatively short shrift while she collected material for a general ethnology of American Samoa, duly published in 1930 as a technical museum monograph, *Social Organization of Manu'a*. This was at most a bit naughty, a bit of a fast one—except that it set her up for the particular episode that is at the center of Freeman's reconstruction. Remarkably, in 1987, Freeman came across one of Mead's principal original informants about adolescent sex, still alive and clearheaded in her 80s. Informed that Mead had told the world about free love in Samoa, and mortified by her role in creating that impression, she swore a deposition to the effect that she and a friend had told Margaret Mead what they thought she wanted to hear. It was a prank not untypical of Samoan humor.

Earlier published versions of this story were greeted by the now-familiar ad hominem: such momentous matters cannot be decided by "octogenarian recollections" (p. 12). Standing alone, the

claim of hoax was just another piece of evidence, subject to standard critical scrutiny. Freeman does his best to test it. There were two separate interviews conducted by Samoan intermediaries; the informant was videotaped, her honesty and religious conscience were invoked by the use of the Bible for swearing. Freeman's direct tests are supplemented, however, by his meticulous reconstruction of Mead's movements. His "theory of the case," if you like, shows how Mead needed the information imparted by her two young women informants because she had not done the surveys and detailed interviewing of a sample of adolescent girls as she should have. She was aware that probing into sexual matters would require a lengthy confidence-building period with each informant, as well as cross checking. But by then Mead was eager to leave the field without taking time to do all this.

Freeman does not delve too far into Mead's reasons for her early departure—a particularly surprising decision, given that she initially contemplated extending her stay or returning. Freeman notes that she found life alone in the field very difficult. She declined to live in a native household on the practical grounds that the open structure would make the solitude necessary for work, not to mention privacy, impossible. She lived instead with fellow Americans but found colonial society stultifying in other ways. Hurricane damage during her stay made matters worse. What emerged was a strong urge to finish as soon as possible and return to her career (and a waiting husband).

There is nothing unworthy in any of this. Mead comes through Freeman's account as serious, spunky, and hardworking. She did collect lots of material very rapidly. She sought explicit permission from Boas to draw conclusions not fully backed by evidence. The most dubious action Freeman records is her deception of the Samoans: she allowed herself to be raised to the status of a ceremonial virgin on at least three occasions. This gave her good access to other young women. But it was gained at the expense of deceiving her hosts about her married and nonvirgin status. (A ceremonial virgin is a virgin with ceremonial status, not someone whose virginity is purely ceremonial.)

So Mead comes out of Freeman's reconstruction as a fine fieldworker whose skill and even brilliance led her to take on too much and to believe that she could complete two projects in less than the time budgeted for one. She thought she had decisive evidence from her female confidantes for her main project, and she rushed it

into print because she believed in the ideas and because she thought she had the correct answer to Boas's challenge.

Even for those readers who find the evidence for an out-and-out hoax difficult to swallow, Freeman's reconstruction of Mead's research progress—from letters, field notes, and diary—shows that Mead did not dig deep enough. Even if what her informants told her was true of their circle at that time, there was need to map its extent, and there was contradictory evidence (the semipublic testing of the virginity of brides; the strong emphasis on premarital chastity; the chaperoning of young women) with which it needed to be reconciled. Making all these concessions, the verdict on whether there was premarital free love in Mead's Samoa is, at a minimum, not proven; at a maximum, highly doubtful.

*The Fateful Hoaxing of Margaret Mead* is clearly and enthrallingly written (at least for those who follow these things) and makes a good case for Boas, Benedict, and Mead's having fostered a decisive wrong turn in American anthropology. Whether there is any hope of rationality and a scientific attitude being reestablished in that politicized and postmodernized field is, however, moot.

I have only one small caveat about *The Fateful Hoaxing of Margaret Mead. Margaret Mead and Samoa* was, *inter alia*, a study in scientific method. In particular, Freeman consciously employed Popper's emphasis on falsification as a way to scientific progress. Refuting Mead created intellectual space for competing hypotheses on the relation of nature and nurture, especially those who view nurture as only one of the determinants of social behavior. Discussion of philosophy of science is almost absent from *The Fateful Hoaxing of Margaret Mead* (though John Ziman and C. S. Pierce are invoked). Whether that was the conscious choice of the author or the wish of the publisher (Westview having replaced Harvard), it is, in my view, a striking absence. Freeman's view in the first of his two books on Mead was correct: a contributing reason why American anthropology took the wrong turn was the hegemony of a false empiricist/verificationist philosophy of science. That false philosophy of science became part of culturalism, and undermining it is part of the project of bringing culturalism down. Freeman's own reception is evidence of this. The shockingly poor level of argumentation among his critics, the inability to distinguish myth from fact, idea from advocate, are typical of what one might call "disappointed positivists": verified facts being unobtainable, they conclude there are no facts. Freeman's critique was treated as an assertion of a different cultural perspective, itself

resting on its own structure of myth. Such intellectual nihilism needs both specific and all-around critique for the benefit of a new generational cohort and the slim hope that it will want to rebuild the subject as a science.

#### REFERENCES

- Freeman, Derek. 1983. *Margaret Mead and Samoa: The Making and Unmaking of an Anthropological Myth*. Cambridge, MA: Harvard University Press.
- Mead, Margaret. 1928. *Coming of age in Samoa*. New York: William Morrow.

## Book Review

---

Dan Sperber, *Explaining Culture: A Naturalistic Approach*. Blackwell, Cambridge, MA, 1996. Pp. 175. \$20.95.

Dan Sperber's *Explaining Culture* is partly an attempt to provide an element, namely, human mental phenomena, which is greatly lacking in contemporary anthropological discussions concerning culture. In the light of recent research by cognitive psychologists in understanding mental phenomena (see *The Adapted Mind* by Jerome Barkow, Leda Cosmides, and John Tooby 1992), Sperber offers an analysis of culture—one that he thinks can no longer be ignored by anthropologists—that articulates the intersection between culture and cognition. Moreover, in an attempt to create a dialogue between both anthropologists and cognitive psychologists, an exchange that is hoped to facilitate a new research program within the social sciences, Sperber presents an interesting and controversial methodology by means of which cultural phenomena can be understood effectively. Sperber upholds an epidemiology of representations (hereafter ER) approach to explaining culture, which is the view that cultural phenomena are mental representations that are widespread due to their contagious effect on human minds. He argues that this ER approach is the study of the millions of microprocesses that lead to the emergence, communication, and transformation of representations associated with intraindividual and interindividual cognitive processes. In Sperber's own words, an explanation of culture is simply an explanation of "how and why some ideas happen to be contagious" (p. 1).

*Explaining Culture* can be divided into three parts. In the first third of the book, Sperber proposes a nonreductionistic brand of materialism, in conjunction with his ER approach, that addresses both the psychological and the ecological elements that comprise cultural phenomena. After making clear the inadequacies of a reductionistic materialist approach to cognition likened to Paul Churchland's *Matter and Consciousness* (1988) and Daniel Dennett's *Intentional Stance* (1987), Sperber begins his analysis by arguing for a materialist approach to cognition that takes seriously the causal efficacy of mental phenomena that are the product of complex material properties (p. 14). In the spirit of John Searle's, *The Rediscovery of the Mind* (1994), Sperber suggests (though not explicitly) that mental phenomena are emergent properties that

---

Special thanks to Michael Bradie and Carrie-Ann Kahn for their comments on an early version of this review.

*Philosophy of the Social Sciences*, Vol. 31 No. 4, December 2001 563-571  
© 2001 Sage Publications

arise from the ensemble of neuronal interaction. Mental phenomena possess causal properties that are then able to affect the system of neurons themselves. It is the top-down causal power of mental phenomena that Sperber takes seriously. This leads him to affirm a version of materialism that he calls "modest materialism" (pp. 12-16). Modest materialism "acknowledges different ontological levels [ontological pluralism] in a wholly material world" (p. 12).

At the other end of the spectrum, Sperber is quick to remind the reader that there is almost no agreement among anthropologists with regard to their study of cultural phenomena because the field of anthropology itself does not seem to acknowledge a theoretical framework—only shared technical terms used to describe particular cultural phenomena (pp. 15-17). Other than a unified resistance to Karl Marx's economic materialism (see *Critique of Political Economy* wherein Marx [1904] asserts that economic [material] conditions determine a group's values and ideas), anthropologists share no theoretical framework from within which a fruitful exchange of ideas is really possible.

However, what Sperber finds disturbing is that most anthropologists, notably defended by David Kaplan (1965) in "The Superorganic: Science or Metaphysics?," think that the theoretical apparatuses used by sociology, psychology, and biology are ineffectual in understanding the concepts of anthropology (e.g., marriage, myth, taboo, totemism, etc.). Sperber reasonably responds that if technical terms are all that anthropologists have, then the ontological status of these terms is ambiguous. In Sperber's own words, "One may acknowledge the expertise of anthropologists in matters cultural, and yet deny that they know (or care) what kinds of cultural things really exist" (p. 16). The answer to this ontological vagueness of technical terms in anthropology, thinks Sperber, is that anthropologists need to realize that their analysis of cultural phenomena (e.g., marriage and myth), which tend to share a host of features that resemble one another, is that of interpretation rather than description. The former has ontological implications that the latter lacks. Interpretations, posits Sperber, are representations that owe their existence to a combination of mental (e.g., beliefs, intentions, and preferences) and public representations (e.g., signals, utterances, and pictures). From this materialist perspective, cultural phenomena are mental representations whose distillation throughout culture/society can be explained through the complex "material interaction between brains, organisms and environment" (pp. 16-26).

After revealing the inadequacies of both reductionistic approaches to psychology and the impoverished state of anthropology more generally, Sperber more fully defends the theme that an ER approach to culture is methodologically the most effective way of explaining cultural representations. Sperber begins by distinguishing three explanatory strategies commonly used by anthropologists: (1) interpretive generalization, (2) structuralist explanation, and (3) functionalist explanation. Interpretive generalization is the strategy of singling out a particular phenomenon in a culture and providing an interpre-

tation of it. This interpretation is then used to explain all similar phenomena in other cultures (see Claude Lévi Strauss, *The Savage Mind* 1966, as paradigm of this strategy.). Anthropologists who employ structuralist explanations argue that there are basic patterns, underlying themes, or simple structures that are common to all cultures. The variation we see in particular cultures is simply a modification of these common patterns or themes (see Patrick Menget, "Time of Birth, Time of Being: The Couvade" 1982, as an exemplar of this kind of explanation). Furthermore, those anthropologists who offer functionalist explanations maintain that an accurate account of cultural phenomena is one that reveals the utility or benefit the cultural phenomena have within the society that produces and maintains them (see Douglas Price-Williams, *Explorations in Cross-Cultural Psychology* 1975, for a discussion on the development of functionalism). As part of his defense, Sperber describes and then rejects these more traditional approaches (pp. 33-49). His basic attack on each of these approaches is that none of them provides the causal factors that adequately explain the emergence and development of cultural representations (note: by "cultural" Sperber means "those representations that are widely and durably distributed in a social group," p. 49).

Having dismissed these traditional strategies, Sperber provides a positive defense of his ER approach. Drawing from his early work with Deirdre Wilson, *Relevance: Communication and Cognition* (1986) on this same topic, Sperber argues that cultural representations have their origin in the transmission of mental representations between individuals. An integral part of this transmission, thinks Sperber, is a cognitive tendency to optimize the "effect-effort ratio." That is, during transmission of mental representations, the content of mental representations becomes transformed and simplified. The newly transformed mental representations, according to Sperber, require less mental effort and provide greater cognitive effects than the previous similar representations. These newly transformed mental representations will most likely be retained within the culture (cultural items) due to their simplicity and cognitive benefits (pp. 52-53). Providing the causal chains of mental representation transmission will yield, according to Sperber, a more robust explanation of cultural representations than do many of the current strategies that anthropologists employ. Indeed, the reader is being drawn by Sperber to surmise that his epidemiological analysis of cultural representations most accurately explains cultural macrophenomena through the cumulative effect of both (1) individual mechanisms that generate and transmute mental representations and (2) interactive mechanisms between individuals that explain the spread and preservation of representations.

In the second third of *Explaining Culture*, Sperber argues that human cognitive processes such as belief and concept formation must be included in any causal explanation of cultural phenomena. Sperber notes that anthropologists need to understand that human cognitive processes, which are either directly (called dispositions) or indirectly (side effects of dispositions called susceptibilities) the product of biological evolution (see Stephen Gould and

Elizabeth Vrba, "Exaptation: A Missing Term in the Science of Form" 1981, and Daniel Dennett's *Darwin's Dangerous Idea* 1995, for a more detailed discussion of this distinction), must be included as a necessary condition in any causal explanation of cultural phenomena (pp. 66-67). The cognitive processes of interest to Sperber are (1) basic concept formation and (2) complex concept formation. The latter includes both (1) metarepresentational ability (e.g., ability to doubt or disbelieve) and (2) evocative ability (the ability to express more fully some partially understood idea of another). Sperber speculates that humans have an innate disposition to develop concepts according to particular schemes of everyday empirical knowledge (e.g., living things tend to be taxonomic, artifacts are characterized in terms of function, etc.) that all languages share (p. 69). On the other hand, complex concept formations, like scientific ideas and religious concepts or institutions, are not so easily grasped and are based on susceptibilities because they were probably acquired as a result of some change in environmental conditions. According to Sperber, then, the human ability to have representations of representations and the ability to provide a better understanding of half-understood ideas at some future date are most likely the product of susceptibilities. Sperber concludes that regardless of how abstract a representation is (e.g., French cuisine), its prevalence can be explained as the product of millions of causally linked microprocesses associated with intraindividual and interindividual cognitive processes.

Still, the reader may want a bit more specificity regarding how individuals process mental representations made public (i.e., cultural beliefs) and how such beliefs are communicated within human populations. Sperber does not disappoint the reader in this regard, although he admits that his answer is rather speculative. Sperber argues that an integration of (1) anthropological speculations on cultural representations and (2) psychological speculations on the cognitive organization of beliefs will provide a provisional solution to explaining cultural beliefs (p. 77; see also Sperber's earlier effort "The Epidemiology of Beliefs" 1990). Regarding (1), Sperber begins by arguing, contrary to some social theorists, that mental representations are more basic than public representations. If one were to follow the causal chain far enough back, it could be shown that all public representations ultimately have their origin in mental representations. Regarding (2), Sperber submits that beliefs represented in the mind can be divided into two kinds: intuitive beliefs and reflective beliefs. Intuitive beliefs (e.g., beliefs about cause, substance, number, etc.) are spontaneous and unconscious beliefs about everyday circumstances. Such beliefs are universal because they owe their origin to intrinsic perceptual and inferential processes. Contrastingly, reflective beliefs (e.g., beliefs about science, religion, myth, etc.) are beliefs that have second-order beliefs about them (pp. 89-92). Sperber speculates that reflective beliefs vary across cultures because the second-order beliefs frequently are influenced by different sources in different environments. Despite the fact that intuitive beliefs are shared across cultures, the diversity of cultural beliefs has its origin in the



sources and modes of transmission of reflective beliefs. Thus, Sperber's synthesis of (1) and (2) leads him to conclude that cultural beliefs can be explained by how they are cognized by individuals and how they are communicated within human populations.

In the last and most speculative part of the book, Sperber draws from Darwinian selection models (and Richard Dawkins's *Extended Phenotype* 1982) to argue that the mind is made up of interacting cognitive modules (genetically specified computational devices), including a single metarepresentational module, both of which are crucial to understanding cultural diversity. Sperber suggests that (1) Darwinian Selection Models (DSMs), (2) Influence Models (IMs), and (3) Attraction Models (AMs) will prove efficacious in understanding that human cultural phenomena are widespread mental representations via public representations. DSMs explain the *human* in "human culture" by pinpointing selected brain mechanisms (i.e., mental modules). Sperber speculates that evolutionary forces are likely to have favored the emergence of specialized and efficient mental mechanisms that are able to take advantage of new information with the aid of old information. The effect is that human cognitive systems are better able to make correct decisions in the struggle for survival. IMs are designed to make clear the term *culture* in "human culture" by pointing out that surviving cultural representations may not be replicas of the cultural representations that preceded them, because the surviving cultural representations may owe their existence to different influences. IMs, which work on the basic premise that influence is a matter of degree, are designed to capture the subtleties of the transmission process from mental representations to public representations. Unfortunately, neither the DSMs nor the IMs fully capture the fact that the human brain creates its own representations—AMs capture this fact better. AMs are statistical models that show patterns of cultural transformations in the direction of some specific point. The trick, Sperber notes, is to isolate this point of attraction and provide a causal explanation that should include both psychological and ecological factors (pp. 113-18). While AMs do not themselves have any explanatory force, they do provide a clearer picture from which an explanation can be procured. (See Elliott Sober's "Models of Cultural Evolution" [1992], Donald Campbell's "Blind Variation and Selective Retention in Creative Thought as in Other Knowledge Processes" [1960], and William Harms's "Cultural Evolution and the Variable Phenotype" [1996], for further discussions on extending the theory of evolution to culture.)

By far the most speculative ideas in *Explaining Culture* are those postulated by Sperber concerning the mind as a complex interacting system of modules. Drawing upon both his own "Modularity of Thought and the Epidemiology of Representations" (1994) and Jerry Fodor's *Modularity of Mind* (1983), Sperber quite nicely sets up the problem with modular views of the mind. He notes that, traditionally, it has been thought that if one views the mind as a bundle of independent genetically specified computational devices (i.e., cognitive modules), then such a postulation cannot account for (1) the fact that

information needs to be integrated for representations to be produced or (2) the existence of cultural diversity and novelty (pp. 120-23). But, contrary to these traditional criticisms, a modular conception of the mind can explain both (1) and (2), thinks Sperber. With regard to (1), cognitive modules are evolved mechanisms that are the result of ancient, gradual, and disordered biological processes. The result is the production of a whole host of connected and disconnected modules performing specific functions. Once a certain level of modular complexity is reached, Sperber conjectures, it is possible for modules to produce additional modules to solve problems internal to the cognitive mechanisms themselves (p. 128). If information needs to be integrated between modules, then existing modules will produce additional modules to facilitate such an integration. With regard to (2), Sperber theorizes further that humans have a metarepresentational module that produces diversity between cultural beliefs. This metarepresentational module processes cultural information and can produce wholly unique reflective representations (pp. 146-50). Of course, Sperber admits that this modular framework of the mind is rather speculative. Nonetheless, he does think that it provides a causal account of cultural diversity that can be confirmed or denied by future cognitive psychologists working in tandem with cultural anthropologists.

While Sperber offers a provocative approach to understanding cultural phenomena that will raise the eyebrows of anthropologists, cognitive psychologists, and philosophers alike, I have a few concerns. First, Sperber would do his readers a favor by clearly defining what he takes to be the difference between *token-token reductionism* (weak reductionism) and *type-type reductionism* (strong reductionism) with regard to cultural phenomena. Considering that cultural types only have epistemological importance rather than ontological standing for Sperber (pp. 21-23), his token-token reductionism and type-type reductionism distinction is not entirely clear. To avoid strong reductionism, types and tokens must both be granted ontological status. Since Sperber only grants ontological status to cultural tokens, he allows himself no room to maneuver away from strong to weak reductionism. Thus, Sperber has left himself with the reductionist label *simpliciter*. This is exactly the appellation that Sperber was trying to rid himself of with the type-token distinction.

Second, Sperber's use of the disposition/susceptibility distinction would be clearer if he provided a good example to illustrate the connection between these terms and how they relate to concept and belief formation. Basically, Sperber considers a susceptibility to be a side effect of a disposition (p. 67). He argues that humans have innate cognitive dispositions that allow them "to develop concepts [or beliefs] according to certain schemas" (p. 69). For example, human concepts for artifacts tend to be categorized in terms of function, human concepts of color tend to be focused on certain hues, and human language-acquisition devices are related to grammar construction (p. 69). These innate cognitive dispositions are the direct product of natural selection, while other mechanisms in the brain (e.g., metarepresentational ability or evocative

ability) are “spin-offs” (i.e., side effects) of these basic dispositions or other more basic dispositions. But this does not accurately capture the biological relationship between these two concepts that are crucial to his project. The following example should help illustrate Sperber’s use of the terms *disposition* and *susceptibility*. Biologists have theorized that the origin of flight in birds is a spin-off from the selective advantage of feathers for thermal regulation. Birds can fly because they became susceptible to flight only as a result of the disposition for feathers. Although Sperber’s entire discussion on the geography of the mind is highly speculative, he should make sure that difficult concepts are grounded in clear examples.

A third concern is with Sperber’s use of a single metarepresentational module. Sperber argues that humans have a single metarepresentational module that “processes concepts of concepts and representations of representations . . . [and] is the set of all representations of which the organism is capable of inferring or otherwise apprehending the existence and content” (p. 147). But it is equally possible that beliefs, desires, and intentions, which are the proper domain of the metarepresentational module, have produced their own metarepresentational modules, respectively. Sperber should leave room for such a possibility. Multiple metarepresentational modules would not hurt the internal structure of his argument in the least.

The last point of contention is with the actual explanatory power behind Sperber’s modularity of mind thesis. He submits that cultural diversity can be explained within a modularity of mind hypothesis. The mind, Sperber speculates, is a bundle of encapsulated modules designed to perform specific functions given particular environmental pressures (p. 133). Not only are these modules able to produce additional modules due to external environmental pressures, but these modules are also able to produce additional modules to help facilitate interaction between modules. Whenever there is an external or internal problem to be solved, then, a new mental module(s) is constructed to resolve the difficulty at hand so long as existing modules are unable to resolve the pressing problem.

On the surface, all this module production is rather convenient. No matter what kind of problem one might present to Sperber about the mind, he could simply invoke the creation of a new module(s). For example, what if I submitted that part of an explanation of dreams requires incorporating the subconscious? Such an explanation, I could argue, could not easily be incorporated into a modularity thesis because the “bridge” that connects the conscious and the subconscious could be the result of an entirely different set of physical relationships that have nothing to do with interacting modules. Sperber could respond that there is a cognitive module(s) in the consciousness that interacts with an existing subconscious module(s). Understanding the processes between these modules, he could say, may help in understanding the intricacies of dreams. My point is that any phenomenon that is presented as a challenge to Sperber’s modularity thesis could be used by him to confirm his theory, thus rendering his analysis of mind unfalsifiable and ad hoc.

Despite some of the concerns noted, Sperber's distinction between the *transformation* of cultural representations and the *replication* of cultural representations is refreshing. He notes that representations that are retained in a culture are rarely the exact same representations that preceded them. Unlike the replication of genetic material, representations are usually transformed by an individual mind that processes information about a given representation and then communicates a slightly altered version to others. In fact, exact replicas of prior representations are limiting cases of transformations. A simple example is the elementary school "telephone game." A group of people are lined up in a room, and the first person is told to whisper a joke to the next person, and so on, until the joke comes back to the original joke teller. The chances of the joke being exactly the same as the original version are very slim. Most likely, the joke has been considerably transformed. In much the same way, Sperber argues that cultural representations get transformed, not replicated, as they are communicated from one person to the next. Sperber's transformation/replication distinction wonderfully pinpoints the limits of the genetic analogy.

In summary, Sperber offers a speculative and highly provocative account of the human mind. This account is grounded in a materialist ontology that pays respect to humans as evolved organisms. Given this framework, widespread cultural phenomena are the accumulated effects of mental representations made public. Sperber has boldly offered this highly speculative analysis of cultural phenomena in the hope of facilitating progress in the empirical sciences through a joint venture between cognitive psychologists and anthropologists. If, at the very least, *Explaining Culture* lays the groundwork for an exchange of ideas between psychologists and anthropologists, then Sperber's work will have accomplished much. It will then be the task of the anthropologists and psychologists to reveal the explanatory power of Sperber's ruminations.

#### REFERENCES

- Barkow, Jerome, Leda Cosmides, and John Tooby. 1992. *The adapted mind*. New York: Oxford University Press.
- Campbell, Donald. 1960. Blind variation and selective retention in creative thought as in other knowledge processes. *Psychological Review* 67:380-400.
- Churchland, Paul. 1988. *Matter and consciousness*. Cambridge, MA: MIT Press.
- Dawkins, Richard. 1982. *Extended phenotype*. Oxford: Oxford University Press.
- Dennett, Daniel. 1987. *The intentional stance*. Cambridge, MA: MIT Press.
- . 1995. *Darwin's dangerous idea*. New York: Simon & Schuster.
- Fodor, Jerry. 1983. *The modularity of mind*. Cambridge, MA: MIT Press.
- Gould, Stephen, and Elizabeth Vrba. 1981. Exaptation: A missing term in the science of form. *Paleobiology* 8 (1): 4-15.

- Harms, William. 1996. Cultural evolution and the variable phenotype. *Biology and Philosophy* 11:357-75.
- Kaplan, David. 1965. The superorganic: Science or metaphysics? *American Anthropologist* 67:958-76.
- Levi Strauss, Claude. 1966. *The savage mind*. Chicago: University of Chicago Press.
- Menget, Patrick. 1982. Time of birth, time of being: The couvade. In *Between belief and transgression*, edited by M. Izard and P. Smith, 193-209. Chicago: Chicago University Press.
- Price-Williams, Douglas. 1975. *Explorations in cross-cultural psychology*. San Francisco: Chandler & Sharp.
- Searle, John. 1994. *The rediscovery of mind*. Cambridge, MA: MIT Press.
- Sober, Elliott. 1992. Models of cultural evolution. In *Trees of life*, edited by P. Griffiths, 17-38. Dordrecht: Kluwer.
- Sperber, Dan. 1990. The epidemiology of beliefs. In *The social psychological study of widespread beliefs*, edited by C. Fraser and G. Gaskell, 25-44. Oxford: Clarendon.
- . 1994. The modularity of thought and the epidemiology of representations. In *Mapping the mind*, edited by L. A. Hirshfeld and S. A. Gelman, 39-67. New York: Cambridge University Press.
- Wilson, Deirdre, and Dan Sperber. 1986. *Relevance: Communication and cognition*. Cambridge, MA: Harvard University Press.

—Mahesh Ananth  
Bowling Green State University

**INDEX**  
**to**  
**PHILOSOPHY OF THE SOCIAL SCIENCES**  
**Volume 31**

**Number 1 (March 2001) pp. 1-136**  
**Number 2 (June 2001) pp. 137-276**  
**Number 3 (September 2001) pp. 277-456**  
**Number 4 (December 2001) pp. 457-580**

**Authors:**

- AGASSI, JOSEPH, "Reply to Professor Gross" [Response], 252.  
ANANTH, MAHESH, "Explaining Culture: A Naturalistic Approach, by Dan Sperber" [Book Review], 563.  
BARNES, BARRY, see Loyal, S.  
BEEBE, JAMES R., "Interpretation and Epistemic Evaluation in Goldman's Descriptive Epistemology," 163.  
BUNGE, MARIO, "Le sens des valeurs, by Raymond Boudon" [Book Review], 122.  
BUNGE, MARIO, "Systems and Emergence, Rationality and Imprecision, Free-Wheeling and Evidence, Science and Ideology: Social Science and Its Philosophy According to van den Berg" [Discussion], 404.  
BURKITT, IAN, "The Later Foucault, edited by Jeremy Moss" [Book Review], 126.  
COLLIN, FINN, "Bunge and Hacking on Constructivism" [Review Essay], 424.  
DOWNES, STEPHEN M., "Agents and Norms in the New Economics of Science," 224.  
DUNBAR, SANDRA, see Risjord, M.  
HANLY, CHARLES, and CHRISTOPHER NICHOLS, "A Disturbance of Psychoanalytic Memory: The Case of John Rickman's Three-Person Psychology," 279.  
HERRERA, C. D., "Research Ethics and the Interpretive Stance in Fieldwork" [Discussion], 239.  
HOCUTT, MAX, see Levin, M.

- JARVIE, I. C., "The Emergence of British Social Anthropology According to George Stocking" [Review Essay], 267.
- JARVIE, I. C., "Freeman on Mead Again" [Review Essay], 557.
- KEITA, L. D., "The Bell Curve and Heredity: A Reply to Hocutt and Levin" [Discussion], 386.
- KEITA, L. D., "Jacobs, Equal Opportunity, and *The Bell Curve*: A Critique" [Discussion], 247.
- KELLY, TERRENCE, "Sociological Not Political: Rawls and the Reconstructive Social Sciences," 3.
- KRIMERMAN, LEONARD, "Participatory Action Research: Should Social Inquiry Be Conducted Democratically?" 60.
- LEMOS, JOHN, "A Defense of Darwinian Accounts of Morality," 361.
- LEVIN, MICHAEL, and MAX HOCUTT, "Reply to Keita" [Response], 395.
- LOYAL, STEVEN, and BARRY BARNES, "'Agency' as a Red Herring in Social Theory," 507.
- MÄKI, USKALI, "Explanatory Unification: Double and Doubtful," 488.
- MARTIN, JOHN LEVI, "On the Limits of Sociological Theory," 187.
- MOLONEY, MARGARET, see Risjord, M.
- MURPHY, DOMINIC, "Hacking's Reconciliation: Putting the Biological and Sociological Together in the Explanation of Mental Illness," 139.
- NICHOLS, CHRISTOPHER, see Hanly, C.
- PEACOCK, MARK, "Taking Wittgenstein Seriously" [Review Essay], 104.
- PICKEL, ANDREAS, "Between Social Science and Social Technology: Toward a Philosophical Foundation for Post-Communist Transformation Studies," 459.
- RISJORD, MARK, MARGARET MOLONEY, and SANDRA DUNBAR, "Methodological Triangulation in Nursing Research," 40.
- ROOT, MICHAEL, "The Problem of Race in Medicine," 20.
- SAUNDERS, B.A.C., and J. VAN BRAKEL, "Rewriting Color" [Discussion], 538.
- STEWART, PETER, "Complexity Theories, Social Theory, and the Question of Social Complexity," 323.
- TRESCH, JOHN, "On Going Native: Thomas Kuhn and Anthropological Method," 302.
- TUCKER, AVIEZER, "Historiographic Realism" [Review Essay], 254.
- VAN BRAKEL, J., see Saunders, B.A.C.
- VAN DEN BERG, AXEL, "The Social Sciences According to Bunge" [Review Essay], 83.
- WALL, EDMUND, "Sexual Harassment and Wrongful Communication" [Discussion], 525.
- WOLFE, JOEL D., "Power of Philosophy" [Review Essay], 111.

### Articles:

- "'Agency' as a Red Herring in Social Theory," Loyal and Barnes, 507.

- "Agents and Norms in the New Economics of Science," Downes, 224.
- "Between Social Science and Social Technology: Toward a Philosophical Foundation for Post-Communist Transformation Studies," Pickel, 459.
- "Complexity Theories, Social Theory, and the Question of Social Complexity," Stewart, 323.
- "A Defense of Darwinian Accounts of Morality," Lemos, 361.
- "A Disturbance of Psychoanalytic Memory: The Case of John Rickman's Three-Person Psychology," Hanly and Nichols, 279.
- "Explanatory Unification: Double and Doubtful," Mäki, 488.
- "Hacking's Reconciliation: Putting the Biological and Sociological Together in the Explanation of Mental Illness," Murphy, 139.
- "Interpretation and Epistemic Evaluation in Goldman's Descriptive Epistemology," Beebe, 163.
- "Methodological Triangulation in Nursing Research," Risjord et al., 40.
- "On Going Native: Thomas Kuhn and Anthropological Method," Tresch, 302.
- "On the Limits of Sociological Theory," Martin, 187.
- "Participatory Action Research: Should Social Inquiry Be Conducted Democratically?" Krimerman, 60.
- "The Problem of Race in Medicine," Root, 20.
- "Sociological Not Political: Rawls and the Reconstructive Social Sciences," Kelly, 3.

#### **Book Reviews:**

- "*Explaining Culture: A Naturalistic Approach*, by Dan Sperber," Ananth, 563.
- "*The Later Foucault*, edited by Jeremy Moss," Burkitt, 126.
- "*Le sens des valeurs*, by Raymond Boudon," Bunge, 122.

#### **Discussions:**

- "The Bell Curve and Heredity: A Reply to Hocutt and Levin," Keita, 386.
- "Jacobs, Equal Opportunity, and *The Bell Curve*: A Critique," Keita, 247.
- "Research Ethics and the Interpretive Stance in Fieldwork," Herrera, 239.
- "Rewriting Color," Saunders and van Brakel, 538.
- "Sexual Harassment and Wrongful Communication," Wall, 525.
- "Systems and Emergence, Rationality and Imprecision, Free-Wheeling and Evidence, Science and Ideology: Social Science and Its Philosophy According to van den Berg," Bunge, 404.

#### **Responses:**

- "Reply to Keita," Levin and Hocutt, 395.
- "Reply to Professor Gross," Agassi, 252.



**Review Essays:**

"Bunge and Hacking on Constructivism," Bunge, 424.

"The Emergence of British Social Anthropology According to George Stocking," Jarvie, 267.

"Freeman on Mead Again," Jarvie, 557.

"Historiographic Realism," Tucker, 254.

"Power of Philosophy," Wolfe, 111.

"The Social Sciences According to Bunge," van den Berg, 83.

"Taking Wittgenstein Seriously," Peacock, 104.

### ERRATUM

In the Table of Contents for the September 2001 issue (Volume 31, Number 3), Finn Collin's review essay was mistakenly identified as "Bunge and Hacking on Constructionism" instead of "Bunge and Hacking on Constructivism" as correctly appeared in text.