

Schering Foundation Workshop 1

BIOSCIENCE \rightleftharpoons SOCIETY

D.J. Roy, B.E. Wynne, R.W. Old, Editors

Report of the Schering Workshop on
BIOSCIENCE \rightleftharpoons SOCIETY
Berlin 1990, November 25–30

Rapporteurs: R.A. Gordon, S.S. Jasanoff, J.D. Moreno, D.W. Sharp

Program Advisory Committee:

D.J. Brahams, H.G. Gassen, R.A. Gordon, D.J. Roy,
R.C. Solomon, W. Van den Daele

Consulting Organizer:

S. Bernhard

John Wiley & Sons
Chichester New York Brisbane Toronto Singapore

Our Brains, Our Selves: Reflections on Neuroethical Questions

Patricia Smith Churchland

*Department of Philosophy
University of California
San Diego, CA 92093, U.S.A.
and
Salk Institute
La Jolla, CA, U.S.A*

Abstract. Research in neuroscience raises a variety of ethical issues, including a) whether humans ought to acquire knowledge about how their brains work, and b) assuming neuroscience does yield such knowledge, what are the ethical limitations on application of that knowledge? The issues are complex, and social policy decisions require understanding of the facts. Religion appears to be neither a consistent guide nor a source of moral understanding. The formulation of sound social policy will depend on nondogmatic and tolerant humans reasoning together. In particular, several common fallacies in reasoning are avoidable.

INTRODUCTION

It is with more trepidation than I care to admit that I have embarked upon this essay. First, because I find ethical issues, unlike their empirical brethren, exasperatingly hard to settle. I am not a utilitarian, because all extant formulations of the utilitarian principle can be counterexamined to absurdity. I do recognize, however, that utilitarian calculations often help in our moral deliberations. I am not a Kantian because the basic Kantian criterion ("moral rules are universalizable rules") either can be grittily adopted by blatantly immoral scoundrels, or the scoundrel can so fine-tune the description of the

rule as to trivialize its application. Moreover, any rule I can make sense of – and apply without cheap tricks – has exceptions that are all too easily dreamt up by any moderately imaginative undergraduate.

Is God perhaps the source of what is right and wrong? Dostoyevsky thought so. But two thousand years before he worried about what is permitted, Socrates had the decisive argument against that theory. He presents it as he questions the student Euthyphro: do the gods approve of something because it is right, or is it right because they approve of it? And Euthyphro, plodding student though he is, comes to realize that the dilemma scotches the gods as the moral fountainhead. For if the first alternative were correct, then standards would be independent of the gods and so the gods cannot be the fountainhead. If the second were correct, then morality would be arbitrary, in that it would be merely a function of the gods' whim, whatever sort of character, nasty or nice, they have. This is a simple destructive dilemma, as the argument form is called in logic. Even if God were to exist, it would not help one iota with moral theory.

It might be expected that a philosopher should have a position on the question of where moral standards come from, i.e., a moral theory, and that the theory would be defended rigorously and vigorously to the hilt. Lest I disappoint, I wish to make a clean breast of things at the outset: I do not have a moral theory, or anything approximating such. I recognize that deep-seated if vague agreements on elementary dispositions concerning justice and fairness must have some basis in our evolutionary history and our genes, but how to make that connection more precise is unknown. Until much more is understood about the brain and what in the wiring is genetically prescribed, the details of the genetic component will probably remain out of reach. Despite these lacunae, we can say that *Homo sapiens* is clearly a social species, like wolves and ants, but unlike orangutans or polar bears, and it is overwhelmingly likely that our genes are an essential element in producing the social disposition characteristic of human brains.

Paul Churchland (1989) is probably correct when he suggests that specific moral concepts are acquired in much the way we acquire concepts for most other categories. That is, our genes give us a headstart with some *in situ* but modifiable neural organization, then the neural networks learn from examples. By error-correcting and/or associating to ever greater consistency with experience, a network constructs representations that cognitively resemble those of its fellows. From example, the child learns what it is to hurt and be hurt, what taking turns, sharing, and exploitation are, and assuming that he is "characterologically competent" (can empathize, can feel sorry, embarrassed, or whatever), he comes to act on his inner representations in social con-

texts, not always perfectly, not always without backsliding and weakness of will but often with remarkable courage and selflessness. At least, that would be the run-of-the-mill case. Moral blindness, on this view, is due to genetic differences, missocialization, or both.

If I do not have a moral theory, how can I presume to go on and talk about morality in the application of neuroscientific knowledge? I shall do it the way we all do it when we are not selfconsciously practicing moral philosophy: we reason together, with tolerance, patience, sympathy and common sense. We draw on such wisdom and empirical knowledge as we may be lucky enough to possess, and we engage in give-and-take. This is crude, I grant you, but I would rather not pretend to more.

There is a second reason why I approach this essay gingerly. The questions we are asked to consider concern social policy on matters that are emotionally inflammable. They seem tremendously important, and they are, yet they become palpable only in tandem with progress in brain science. This means that there are frequently too few facts to sustain a considered judgment. I fear, therefore, that I shall often flounder, and consequently I shall be gratified if even a small amount of light can be shed upon these questions.

What follows is my perspective on some questions concerning the possible applications of neuroscientific knowledge. I must confess, however, that it leaves out much, and I sometimes skate rather fast over the ice – especially the thin bits. What I shall do is sketch a kind of logical geography of the major features in the landscape, as I see it, in hopes that this will provide a relatively clear basis for further discussion. In this survey of the terrain, I have tried to include not only matters that exercise me, but also some that are of greater concern to others. They are not all equally compelling or profitably discussable.

A wide range of morally relevant issues concerning neuroscience can be listed, and for convenience, I divide them into two broad classes: 1) those that pertain to the just and proper application of knowledge, and 2) those that focus on the propriety of having knowledge about brain mechanisms, quite apart from any practical application the knowledge might have. In the first category, we shall consider a) *potential gains* (diagnostic, preventative and therapeutic) and b) *potential abuses* (clinical, political, judicial, and experimental). In the second category, we shall consider a) the "*Faustian potential*", as I call it; that is, the possible loss of virtue or "humaness" that neuroscientific knowledge may visit on us, because it is "knowledge we should not have"; and b) the experimental cost.

I shall address these two classes of issue in reverse order.

IS NEUROSCIENTIFIC KNOWLEDGE WORTH HAVING?

The "Faustian potential"

Should we know how our brains work? Progress in understanding the world has often been viewed with suspicion and disgust, and even now scientists are not infrequently portrayed as unscrupulous, antisocial, and impractical, as well as having a dose of madness and a fondness for the darker arts. There is a long-standing tradition that objects to accretions in empirical knowledge, usually because in some way the discoveries are thought to entail a loss of something special, something spiritual, or something "essentially human". In this instance, the view, as I reconstruct it, is that, as we come to understand the nature of the mind-brain, it entails that we, as human individuals, are demystified, and hence that we are dehumanized, demoralized, and dissected like planaria. Falling in love is a favorite example of what, apparently, we ought to prefer unexplained, for, the argument goes, should we explain it, love would become just another physical process, and a process that might even be duplicatable in a robot. The excitement, the mystery, the abandon that is love would be as lost in the dissection, even as life is lost in a dissected laboratory specimen.

This point of view is best opposed by a liberal education, for no single argument or two will have much force against a generalized conviction that ignorance is to be preferred to knowledge. Although the idea that knowing less is understanding more may be a consolation to the ignorant, it is more lamentable than anything else. In any case, it is gratifying to note that in the hurly burly of life, the dictum may be mouthed but not lived.

Unlike my college freshman class, this audience does not need perorations on the practical value *and* aesthetic virtues of empirical knowledge. Apart from the genuine aesthetic pleasures resulting sheerly from solving the mysteries, and apart from the glaring fallacy in equating understanding to killing-by-dissection, the medical and creature-comfort benefits deriving from science are obviously stupendous. Perhaps because the heyday of torturing the insane to drive out devils is long past, because we scarcely remember smallpox, polio, or undulant fever, and because we are long past the days when the Catholic church condemned anesthetics and vaccines as instruments of Satan (White 1896), and because antibiotics have demoted pneumonia from a main event to a sideshow, the technological spin-off from science is taken for granted. The debt to science as a headwater for practical benefits is very easy to neglect.

My further personal perspective on the Blakeian dogma that knowledge entails loss of a precious spiritual innocence and that the loss outweighs the gain, is that, by and large, it just does not make sense. Some chord or other must have been struck by Blake, and perhaps I am just lamentably tone deaf. But I have never seen the argument, only the poem¹; I have never heard the justification, only the rhetoric. As a farm child struggling to help make a living in a world where practical benefits meant the difference between food and hunger, or between machines and hand-toil, I long believed the choice to be dead obvious.

Nevertheless, because the arguments that champion ignorance over scientific understanding have a sympathetic following both in so-called "New Age" groups as well as in certain segments of organized religion and the entertainment industry², they are disregarded at our peril. They do have an effect.

What price knowledge?

Animal experimentation. The main questions under this rubric have concerned whether humans ought to test hypotheses on animals, and if so, which ones and under what conditions. Although computer modeling is playing a growing role in neuroscience, there is no doubt that it cannot replace animal experiments. If neuroscience is to continue, animal experiments are necessary. In the history of biomedical discoveries, animal experimentation played a crucial role. Banting and Best managed to isolate insulin, but only with the aid of live animals; Salk's discovery of polio vaccine, and the development of birth control drugs, depended essentially on knowledge gleaned from experiments on animals. Antivivisection groups are sometimes opposed to all animal experimentation, on whatever kind of animal, and for whatever purpose, whether it is testing AIDS vaccine in monkeys, or contraception in mice, or locomotion in cockroaches.

In the U.S.A, the Society for Neuroscience, especially under the leadership of David Hubel from Harvard and Patricia Goldman-Rakic from Yale, has begun to organize responses to the antivivisectionist campaign. In August of 1990, The British Association for the Advancement of Science launched a

1 William Blake, "The Songs of Innocence and Experience". In: Keynes (1966).

2 For a wonderful if blood-souring sampler, see Martin Gardner's *The New Age: Notes of a Fringe Watcher* (Gardner 1988). This includes discussions of the views of Margaret Mead, Shirley MacLaine, Arthur Koestler, Oral Roberts, Pat Robertson, Jerry Falwell, L. Ron Hubbard, Russell Targ, Harold Puthoff, and many others.

declaration concerning animals in medical research, which is balanced, principled, and sane. It forthrightly articulates the need for animal experimentation in research and for scientists to explain this to the public at large.³ The International Brain Research Organization (IBRO) has established a committee⁴ to address the threats posed by animal liberation groups in many different countries.

These developments strike me as extremely important, since academics are more inclined to hope the problem will go away than to respond politically, and because widespread, hideous diseases such as schizophrenia and Alzheimer's disease cannot be brought under control in the absence of animal experimentation. Barring a new target to siphon off the antivivisectionist fervor, I gloomily forecast that the interference with medical research will continue unabated. Although the antivivisectionists may have done some good in requesting adherence to standards by cosmetic companies and abattoirs, they have also done great damage to medical research programs and to animal study programs that, paradoxically for the animal liberationists, provide results that precisely benefit animals and promote the humane treatment of animals. Research on pain has virtually ground to a halt, because scientists cannot undertake the risk of testing drugs on animals given experimental pain. The number of humans and animals suffering pain, caused by many forms of cancer, spinal injuries, burns, and so forth, has not, however, declined. So far as the moral principle is concerned, it is certainly not obvious to me that no

- ³ See *IBRO News*, 18(3): 1. The text reads as follows:
 In view of the threat to medical research posed by increasingly vocal and violent campaigns for the abolition of animal experimentation, we make the following declaration:
 a) Experiments on animals have made an important contribution to advances in medicine and surgery, which have brought major improvements in the health of human beings and animals.
 b) Much basic research on physiological, pathological and therapeutic processes still requires animal experimentation. Such research has provided and continues to provide the essential foundation for improvements in medical and veterinary knowledge, education, and practice.
 c) The scientific and medical community has a duty to explain the aims and methods of its research, and to disseminate information about the benefits derived from animal experimentation.
 d) The comprehensive legislation governing the use of animals in scientific procedures must be strictly adhered to. Those involved must respect animal life, using animals only when essential and as humanely as possible, and they should adopt alternative methods as soon as they are proved to be reliable.
 e) Freedom of opinion and discussion on this subject must be safeguarded, but violent attacks on people and property, hostile campaigns against individual scientists, and the use of distorted, inaccurate or misleading evidence should be publicly condemned.
- ⁴ *Ibid.* pp. 1 and 6.

rat should be caused pain in order to relieve a child from suffering terrible oncological pain.

Are scientific research animals in fact poorly treated? This is a very broad question, but let me say that, in my experience, the answer is emphatically "no". The regulations governing animal research are actually very strict, and in California, for example, laboratories are routinely inspected, unannounced, by county, state, and federal officials. Typically, researchers care a great deal about the comfort and health of their animals, both for reasons of human decency and because sick or stressed animals do not give reliable results.

It is fairly obvious that if the antivivisection campaign were successful, much of neuroscience and other biomedical research would be impossible. Assuming that human suffering has some place in the calculation of whether this would be acceptable, I can but conclude that animal experimentation is an issue where moral fervor may fail, and fail egregiously, to coincide with moral probity and moral decency. It may also be worth noting that in San Diego, for example, the number of stray dogs put to sleep by the Humane Society is two orders of magnitude greater than the number of dogs transferred to laboratories. My guess is that these figures are not atypical of other major cities.

Human experimentation. i) *Studies using invasive techniques.* These are of course limited in the extreme, but under careful conditions, some data can be gathered in the course of surgery or chronic electrode implant. For example, George Ojemann, a neurosurgeon in Seattle, has done outstanding work this way (Ojemann 1983; Ojemann et al. 1989). Ethics committees watch this sort of research with great care, as do granting agencies and the journals to which the research papers are sent, and Ojemann himself is a clinician first, and a researcher second, by which I mean that he puts the interest of the patient ahead of research goals. With the advent of new techniques, the human subjects committees will expand their province.

ii) *Studies using noninvasive techniques.* Although these are generally preferred to invasive techniques from an ethical point of view, they too can be hazardous. Positron emission tomography (PET) uses radioactive material, albeit in tiny amounts, and while magnetic resonance imaging (MRI) appears to be safe, no one knows for sure what the subtle effects might be of putting a brain in a strong magnetic field. Electroencephalography (EEG) is probably the safest technique, but because of problems in spatial resolution, it is best used in conjunction with other techniques. So it is not a replacement for the other techniques.

iii) *Use of Foetal tissue.* Tissue grafts in damaged brains may be an important treatment, especially for Parkinson's disease, and conceivably for Alzheimer's disease. Although the research is still in its infancy, and it is not yet clear what exactly is accomplished by the implanted cells, the results so far (Bjorklund et al. 1990) are mildly encouraging. It goes without saying that the possibility of treating Alzheimer's and Parkinson's diseases has sparked widespread humanitarian interest. There is also much to be learned about neural development and developmental disorders by research on foetal tissue. Here, also, the potential benefits are immense.

The objections to research on foetal tissue derive mainly from anti-abortionists. From their point of view, abortion is murder, and foetal research simply compounds the felony. On the other side, it is argued that so long as abortion is legal – and in the U.S.A and elsewhere it is – destroying tissue which could benefit the suffering is irresponsible. Consider a slightly different case. Jehovah's Witnesses regard blood transfusions as deeply sinful, yet the medical practice is perfectly legal. Government restrictions to prevent transfusions, on grounds that the Jehova's Witnesses are convinced of its sinfulness, would generally be considered an attempt to foist a personal religious dogma on one and all. Does not the same response apply to those whose personal religious convictions deem abortion sinful and the use of foetal tissue heinous profiteering?

My own view is that it is as unjustifiable to forbid early abortion of a foetus as to forbid blood transfusions, or anesthesia or consumption of pork or driving of cars by women, or cremation, or any one of a dozen other by-laws of some particular religion. One is free to follow these strictures oneself for one's own peculiar reasons, but there is no justification for imposing them on others who decline to submit to the creed. My view on the permissibility of abortion coincides with the 1973 decision of the U.S. Supreme Court in the case *Roe vs. Wade*, which upheld a woman's right to an abortion, and to a similar decision of the Canadian Supreme Court in 1989.⁵ Since this is not the occasion for a searching discussion of that issue, I conclude by noting that when different and substantial constituencies of sane, decent, kind, and thoughtful humans disagree about the moral permissibility of an action, then tolerance and mutual respect, not self-righteousness, coercion, and intolerance, are the virtues that ought to prevail. Moral conviction is a psychological state with a welter of epigenetic origins; it is not, as human history shows all too clearly, a singular, let alone an infallible, guide to moral recitude.

⁵ See Kristin Luker's *Abortion and the Politics of Motherhood* (Luker 1984).

Regulating use of foetal tissue is of course desirable, as is regulation of kidney, heart, and blood donations. Indeed, when it is permitted, it is regulated. To prevent abuse of donor-recipient interactions, research hospitals have typically adopted anonymity policies, which keep the two parties and their doctors quite separate, and which require consent of the pregnant woman to the abortion and the use of the tissue.

Did the patient consent? The matter of informed consent is not new and it is virtually always very thorny. It is even more thorny in the case of diseases of the brain, because the patient may be cognitively disabled by the disease. The disease itself may prevent the patient from giving his consent to treatment, as, for example, in alcoholism and other addictions. This matter has already concerned psychiatry for some time, independently of recent discoveries in neuroscience. Consequently, ethics committees are very stringent in evaluating proposals for treatment without consent, as well as for clinical trials and human epidemiological studies. So there is a foundation of policy in place on which to build. Nevertheless, the possibility of abuse remains very real, either because the patient was improperly treated, or because he was *untreated*, owing to the clinician's fear of malpractice litigation. (The question of coercive therapy will arise in part iv in the section "The Ethics of Knowledge Application")

How much will it cost? Research is expensive, and modern research in neuroscience, though cheaper than particle physics, generally requires a heavy research investment. For example, the start up cost for a PET scanner is estimated to be roughly seven or eight million dollars. Operating costs are added on thereafter. Some societies may conclude that they cannot afford high technology neuroscience, and either depend on other nations to do the research, or agree to continue at their current level of medical practice. Funds for neuroscience research in the U.S. are currently very tight as the funds for some government agencies are either cut, or fail to keep pace with increasing costs and increasing numbers of researchers. The costs of research have to be balanced against the costs of other projects society values, such as savings and loan rescues in the U.S. (at least 500 billion).

Against the presumption that neuroscience will have only biomedical benefits as the payoff on the investment, we must pause to acknowledge that pure research often yields discoveries that lead to unforeseen but monumental technological breakthroughs. It is entirely probable that discoveries in neuroscience will fit the mould. Indeed, it has already been conjectured that the brain style information processing currently under study may lead to a revolu-

tion in computer technology, with profound economic sequelae (Caudill and Butler 1990; Hecht-Neilson 1990; Churchland and Sejnowski 1991). Given extant network machines, this seems quite likely, and some devices, such as a network for detecting plastique hidden in suitcases at airport checkpoints, are already functioning.

Politicians regularly benefit from a gentle instruction regarding the value of pure research and its unpredictable practical advantages. The investment in computer research in the sixties or into properties of substances at supercold temperatures might have seemed unwise – might have been “Proxmired”⁶ – at the time, but with the development of chips, electronics, and superconductivity, the investment reaped a bonanza.

THE ETHICS OF KNOWLEDGE APPLICATION

Educated guesses and wild speculation

What knowledge will neuroscience bring? Could we apply it to halt the progress of Alzheimer’s disease, or “rewire” a psychopath, or control thoughts and emotions by “brain bediddlers”? Although some of the speculations can be amusing, my own preference is to concentrate on questions of knowledge application in the light of what knowledge is actually available now, and what we can reasonably expect to know in the foreseeable future. Uninformed speculation can be an awful time-waster, as, for example, in working out the details of irrigating from the canals on Mars before the properties of the planet were more clearly apprehended. There is no sharp line, of course, between educated speculation and educated-but-slightly-nutty wool gathering. For the most part, however, I prefer to think of myself as staying mostly with the first, but I allow that from time to time I may have strayed into the second.

Intervention and exploitation

One major focus of concern as knowledge of the brain progresses is the potential for abuse. Important advances in clinical interventions that treat brain disorders, both psychiatric and neurological, have made impressive progress

⁶ William Proxmire is a Democrat senator from the state of Wisconsin, who became well known for his “Golden Fleece Award”, which sometimes included scientific projects whose motivation and significance he was unable to understand, as well as military costs that would have been hilarious but for the impoverished social programs.

since the 1950s. The medical opportunities for therapy are based on research in a number of fields, but especially neuroscience and molecular biology. Psychoactive drugs, neurosurgical procedures, and genetic analyses that reveal markers for disorders such as trisomy 21 (Down syndrome) or Huntington's chorea are prominent in the roster of remedies. Based on their current capacities and future potential, one might also speculate about further developments in pharmacology and genetics, and perhaps even wonder whether "wiring pathologies" might someday be revealed by high resolution scanning machines.

There is a plethora of moral issues here, ranging from whether antidepressants prevent some persons from confronting and changing their truly depressing situations, to whether one has a duty to abort a foetus carrying a gene for a disease certain to cause dreadful suffering. My own view is that each sort of problem has to be dealt with in the context of the facts pertaining to it – the pharmacological, genetic, diagnostic, epidemiological, etiological, and sociological data. They cannot be satisfactorily discussed all of a bunch, since answers to one problem may be tangential to another. Nevertheless, certain general principles and misconceptions can be efficiently introduced early in the conversation. I have taken the liberty of assuming that others, more qualified than I, will present various problems in their specific settings. My focus will instead be on some abstract and general questions concerning the dangers inherent in using our knowledge of how the brain works to alter human brains.

i) Not my good but *their* good. In its most general formulation, the worry is this: therapeutic techniques for treating damaged and disordered brains will be used for control and exploitation – not only of diseased patients, but of perfectly normal humans so that they become pawns or slaves of someone else. The envisaged scenario is very different from enslavement documented elsewhere in history, for here the brain might be altered to change the person's desires, motives, feelings, and the very way he reasons and plans, and even his conscience about right and wrong. The very possibility of slave revolts would be abolished from within, by making the slave brains compliant and content brains, by making the slaves happy slaves.

This is not a far-fetched danger. The heydays of lobotomies (Valenstein 1986) and valium and electroconvulsive therapy (ECT) for social malcontents have provided a harrowing glimpse into the effects of unscrupulous medical (mal)practice. What can be done to forestall the abuse of neuroscience?

My response here is not very glamorous, but it is the only one that makes sense to me: continue trying to regulate use of techniques, to educate widely,

and to pressure the politicians and clinicians, so that we reduce the abuse to the minimum possible. Abuse of knowledge is of course something for which society must always be vigilant, but for which general rules – other than rules with a hollow ring, such as “do not abuse knowledge” – are not forthcoming.

Undoubtedly the early kindling of fire led to using fire as a weapon and a means of power; the discovery of anesthetics led to withholding in exchange for compliance; harassing phone calls are a menace, and so on *ad nauseam*. If we merely sermonize, “Use knowledge wisely and do not abuse it”, that is not going to deter those who convince themselves they are wise and are making good use of their knowledge. More generally, we cannot expect that fair and decent rules will successfully impede determined and devious rule breakers. Not in medicine, or football, or politics, or anywhere else. We do the best we can within the social reality we start with. Specific rules, regulating where fires can be lit, who can own guns, and the penalties for phone abuse are, on the other hand, reasonable. The updating of regulations and the introduction of new ones, as new technology becomes available, is a never-ending process.

In view of the potential for harmful application of neuroscience, a “hands *completely* off” policy may be urged. According to this argument, the risk is such that we should collectively abjure any sort of medical intervention in human brains. The implications of this recommendation bear long reflection. About 1% of the population is schizophrenic, 12% of men and 18% of women have a depressive disorder, 2% have epilepsy, 20% of people over 65 have Parkinson’s disease, 12% over 60 have Alzheimer’s disease, 10% of American are alcoholics, and so on. What of tumors, gunshot wounds, encephalitis, meningitis, spinal cord paralysis, hydrocephalus, and so forth?

Presumably the “hands off” argument rests on some sort of cost-benefit analysis, but in the absence of having it to hand to study, one can only suspect that a decimal point must have strayed far from where it should be. In any case, it is sometimes forgotten that failing to act can be every bit as immoral as an overt act. Sins of commission may be more colorful, but they are not necessarily more heinous, than sins of omission. Throwing an infant overboard is morally unacceptable; so is failing to pull a drowning infant out of a puddle because one does not want to muddy one’s shoes. Dosing a rambunctious child with ritalin, because we would rather watch TV, is morally unacceptable; letting a hyperactive child become a social pariah, because someone else might use ritalin selfishly, is also unacceptable. We cannot keep our hands clean and our consciences clear by omitting altogether to apply neuroscientific knowledge. What do we do then? The homespun wisdom here is that we muck on as decently well we can, avoiding extremes, and relying on those with proven character, breadth of experience, and depth

of understanding to lead the way. In a more literary vein, one might say that between the Scylla of abuse and the Charybdis of abolition is the painstaking and thoughtful working out of social policy.

ii) The fallacy of the beard. Sometimes in the debate over social policy, it will be noted that clean-cut divisions, between what should count as within the rule and what not, are hard to make. When legislation lays on a division even so, it may seem arbitrary and unjustified. Reflecting on when it is permissible to abort a foetus, some might suggest that drawing the line at five months is arbitrary – after all, there are minute differences between a foetus at 150 days and at 151 or 149 days. And there are many more arguments cut to the same pattern. Why, it might be argued, should the freeway speed limit be set at 55 mph, rather than 56 or 59? Why should garden watering be permitted in San Diego between 10 pm and 8 am, rather than 9:30 pm and 8:30 am?

Observing that there are continua is commonplace; the fallacy consists in concluding that therefore no meaningful distinctions can be drawn. The fallacy of the beard takes its name from the case where it is noticed that there is an uninterrupted continuum between the face of a man who is smooth shaven, the same man at 5 o'clock later that day with the legendary "shadow", the same man next morning with stubble, and finally, the same man a month later with a full beard. To deny that we can draw any distinction between being clean shaven and having a beard, on grounds that any two adjacent points on the continuum are scarcely distinguishable, runs foul of common sense. Large rivers are easily distinguished from small creeks, though some creeks at flood are larger than some small rivers in high summer. But so what? The fuzzy intermediate cases do not negate the contrasts in distant points: hills and mountains, blue and aquamarine, juvenile and adult, manslaughter and negligence, aggressive marketing and price gouging, education and propaganda, conceptus and infant.

In a continuum, the end points typically contrast vividly. A neonate is radically different from a conceptus; a submachine gun is very different from a slingshot; a manic depressive from a blue and lonely first year student. Consequently, when practical life demands it, we can make distinctions on the continuum, recognizing that adjacent points may be virtually indistinguishable. How do we do that? There is no generally applicable answer, save that in moral matters it is wisest to err on the side of caution. Thus, legislating five months rather than seven months as the limiting date for abortion errs, if it errs at all, on the side of caution. So far as distinctions relevant to applications of neuroscience are concerned, here, too, we shall be guided by

available information, moral willingness, and the rule of thumb to err, if at all, on the side of caution.

iii) The Rifkin fallacy. Another temptation in this arena consists in 1) imagining future possibilities based on freely speculative and grossly under-defined technology, 2) discovering that no clear answer about social policy is reachable in these conditions, and then 3) generalizing to conclude that moral matters arising from currently planned and quite well-defined technology are also beyond our collective wisdom. The corollary is that if we do not know how to solve the moral problems served up by a technology, we have no right to use the technology.

“Suppose”, a student will ask, “that neuroscience allows us to produce a device for rewiring sociopaths so that they became socially responsible, but the side-effect is that their mature height is three inches shorter than it would otherwise be, or that they become musically insensitive. Would that be morally acceptable?” The instructor heaves a sigh. When, as in this example, the conditions are woefully underdescribed, who on earth knows *what* to say? To generalize from a freshly hatched “what-if” story to a fledgling medical application is evidently not justified. Factually rounded-out examples are very different from fictional line drawings, and, in particular the richness of the available data in the former case enables us to sink our collective teeth in and figure out a fair-minded course of action. By way of contrast to the fanciful rewiring story, consider more palpable questions such as the use of foetal grafts in Parkinson’s patients, or whether to abort a foetus carrying a genetic marker indicating it is probably a psychopath, or whether, since an hour of clinician’s time is worth about two months of fluoxetine capsules, psychotherapy should be paid for from the patient’s purse, not the purse of the insurance company or the government.

I should perhaps end this section by explaining that I refer to the inference from a fanciful case to realistic cases as “The Rifkin Fallacy” because at least some of Jeremy Rifkin’s objections to scientific research and biomedical innovation have that character. Some, though probably not all, of his acolytes too have acquired a reputation for sliding into the fallacy in the name of public good.

A crude graph with a “trait undesirability” continuum on the X axis and a “biological knowledge availability” continuum on the Y axis, can be constructed, thereby plotting out various genetic counselling issues as a function of what we know and how seriously debilitating is the gene. (Fig. 1) Thus cystic fibrosis and trisomy 21 and trisomy 18 are located approximately in the upper-left, whereas gender is located in the upper-right quadrant. But

traits such as knock knees and IQ are somewhere in the lower-right quadrant. Now, what I am calling fallacious is the idea that arguments pertaining to the acceptability of genetic choice of traits in the lower-right applies equally to genetic choice of traits in the upper-left. The Rifkin fallacy in this arena consists in exactly that. Thus it may seem convincing that one is not justified in aborting a fetus that merely has an average IQ, in hopes that the next fertilization will produce a specimen with a higher IQ. But even should the argument be convincing for traits in the lower-right quadrant, it cannot be neatly rolled over into an argument against aborting a fetus with traits in the upper-left quadrant, such as trisomy 21.

iv) Responsibility: Full, diminished, and absent. Determining social policy for violations of laws concerning public safety is often anguished; it is, however, a domain where practical decisions are unavoidable. In the forefront of unacceptable behavior are violence (including murder, rape, assault and manslaughter), theft, fraud, kidnapping, bribery and tax evasion. Listed among the issues concerning punishment and blame are these: 1) protection of others, 2) deterrence of potential violators, 3) prevention of recurrence by the same person, 4) compensation to the victim, and possibly 5) retribution. A crucial concern in deciding how to treat the person whose behavior is problematic is whether he or she was responsible for what was done. There is a contrast between someone who, during an epileptic seizure, strangled a seatmate, and a young man who plans and carries out his wife's death, leaving him rich and free to marry his mistress.

When is someone responsible for what he does? As more is understood about behavior and its causes, the more complex become the answers. My approach here will be to outline a framework for discussion by focusing on 3) above, namely, prevention of recurrence by the same person. One view of punishment is that *inter alia* it serves as negative reinforcement, and hence deters recurrence, but only if the agent was genuinely responsible for what he did. Moreover, it is generally agreed that it would not be just to punish someone for something that was not, strictly speaking, his fault. While we can distinguish cases of sheer accident from cases of plotting by a sane man, there are many cases where the issue is less clear.

Sometimes the brain has properties which entail that the person has diminished responsibility for the action. For example, the person is paranoid and believes resolutely that the postman is Himmler, and, in the name of the public good, does him in. There are cases where the mechanisms for motor paralysis during REM sleep are defective, and the sleeper acts out his dream of strangling a mad dog, killing his wife as he does so. In these sorts of cases,

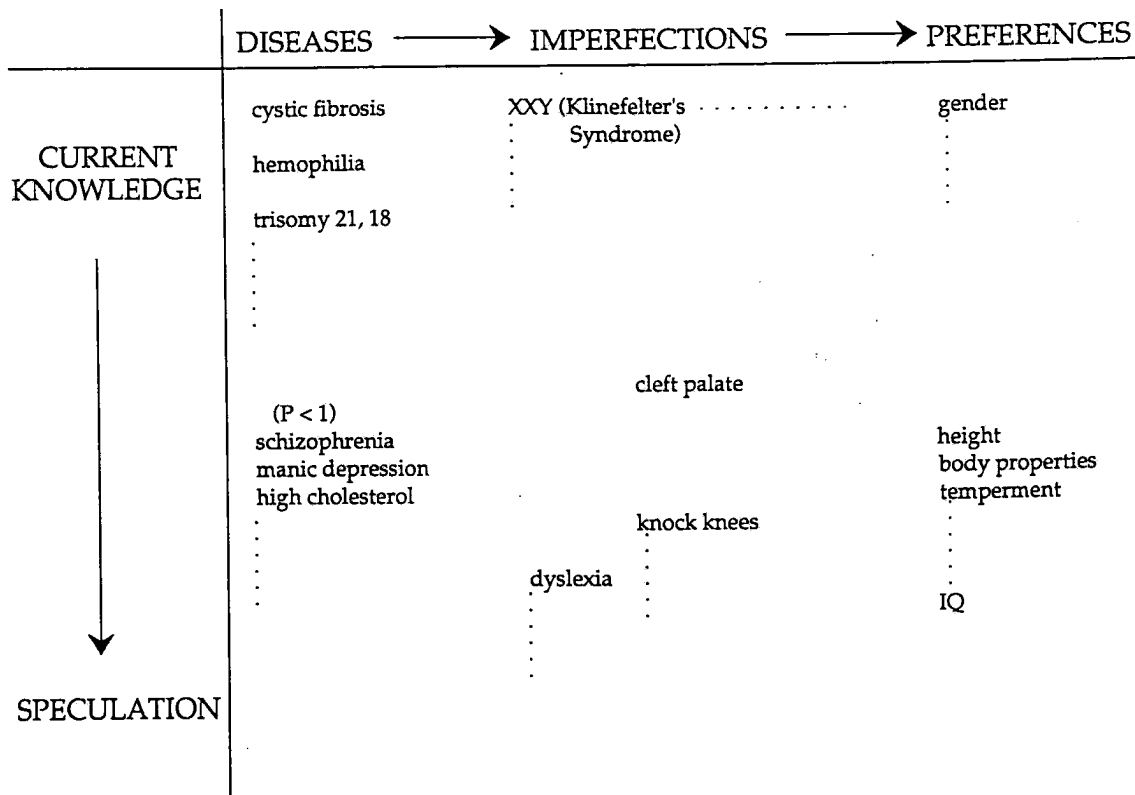


Fig. 1.

punishment may be inefficacious and inhumane. Instead, it may be preferable to intervene in the brain directly. That is easy enough to do in the case of the dreamer, since REM blockers can be used with great success. The person is no longer a threat, but he also no longer dreams: a fair exchange, one would assume.

The paranoid schizophrenic is also treatable with drugs, as is the epileptic. Protective custody may still be in order, but a conventional prison term would probably not. Obsessive-compulsive disorder is now successfully treated with drugs in many cases, and “uncontrollable rage”, manic depression, and erratic acts associated with premenstrual syndrome (PMS) are also argued in court as more appropriately treated with biomedical intervention such as surgery and drugs than with conventional punishment. Genetic markers for some disorders such as schizophrenia and sociopathy may be found, raising the further question whether preventative treatment would be justifiable. If we had known Ted Bundy was a homicidal sociopath before he killed several scores of people, would it have been appropriate to treat the condition with drugs or surgery? How much different would life in Uganda have been with Idi Amin on haldol?

So far we have considered mainly the needs of society, given the deviant behavior of an individual, but we must balance these against the needs of the individual. Privacy, diversity of personality types, eccentricity, and so forth are not features we want to crush on behalf of public protection. Fanaticism about the latter can lead to injustice on a massive scale.

In character, the problem is really one of constraint satisfaction: there are many important constraints – protection of society, and freedom of the individual, to mention only two – where some constraints may be weighted more or less than others. These are tough problems, not to be solved by the mechanical application of set of a hide-bound rules. Constraint satisfaction problems can be elegantly solved by connectionist (neural net) models, given feedback corrections by a “teacher” who knows the answer.⁷ The human condition is such that we do not have anything really analogous to the connectionist “teacher”. The real world with its uncompromising causal reactions is a kind of teacher, but typically an unforgiving one. Nevertheless, we do have each other, and we do have our own neural networks in our heads. Perhaps, reasoning together, continually reassessing and revamping, using gentle persuasion, is the way to approach our constraint satisfaction problem with

⁷ For an introductory account of the computing style of neural nets, see Paul M. Churchland's *The Neurocomputational Perspective* (Churchland 1989) and *Matter and Consciousness* (Churchland 1984); or Patricia Churchland and Terrence Sejnowski's *The Net Effect: A Computational Neuroscience Primer* (Churchland and Sejnowski 1991).

neural networks. As always, it is a mistake to suppose that the issues can be settled on "purely moral grounds alone" and in the absence of relevant scientific data. Such data provide crucial additional constraints.

The adversarial system of courtroom justice (prosecution, defence judge and/or jury) appears to work better than anything else, but sometimes, as we all know, justice is not done, and there is no "Court of Angels" to swoop down and do the job properly. Sometimes we are overwhelmed by the problem of constraint satisfaction, sometimes, as with such problems generally, there may be no one right answer, only a set of "local minima", and sometimes, because we are ignorant or frightened, we get stuck in a local minimum,⁷ and it takes a shock to bounce us out.

v) Will neuroscience eventually show that no one is responsible for anything? My considered view is that, as we come to understand in greater detail how the brain works, we will deepen and broaden both our conception of agency and of the fair means to intervene when there are defects of agency. But a wholesale dumping of the notions of agency and responsibility is not something forseen from where we are in 1991. At least because it may turn out that "normal" individuals are at their best when they conceive of themselves as responsible agents, and because education and nurturing may proceed most effectively if children are taught to conduct themselves as rational, responsible agents. On the other hand, it is also conceivable that new concepts – neuroscientifically informed concepts – may function even better in these roles, and that, with developments in neuroscience, they will evolve and come to replace the orthodox conceptions of agency and responsibility. What such notions could look like is not predictable at this stage of science, at least not by me. But it is worth allowing for the possibility that we can do better than we are.⁸

CONCLUDING COMMENTS

The possibilities before us for understanding and treating human brains are both inspiring and sobering. While there is a clear potential for the abuse of knowledge – as there is in every field of human affairs – vigilance, goodwill, and common sense can do a great deal to steer us around the hazards. I forsee the greater dangers to lie not in new knowledge and its regulated

⁸ See my *Neurophilosophy* (Churchland 1986) or Paul M. Churchland's *The Neurocomputational Perspective* (Churchland 1989).

application, but in fanaticism, superstition, and ignorance, both moral and empirical. I am probably more afraid of the moral fanatic, the self-righteous inquisitor, and the hot-blooded do-gooder, than I am of the self-interested hedonist, who often can be counted upon at least to be reasonable, if not altruistic. I expect we have more to fear from overpopulation than from neural transplants, from holy wars than from genetic engineering. Despots who control us from afar by cognitive/affective "brain beddilling machines" are a science fiction that divert us from the living, breathing despots who control by mob hysteria, advertising, chauvinism, sexism, and religion, none of which require sophisticated neuroscience. This is not to detract from the real neuroethical questions that need to be addressed, but only to suggest that we put them in perspective within the wider range of moral issues that involve changing the brain.

Aristotle believed in moral progress. In his view, as we search and reason about social life and its perils, as we experience life, and reflect on its complexities and surprises, we come to a finer appreciation of what is decent and fair. We learn from each other, and from those whose lives exemplify the human virtues. We learn from errors of the past; our own, and those in the history of the species. This is not a flashy theory of the archetypal Good, nor is it the sort of theory to whip up moral zeal. Nevertheless, it is a reasonable and sensible approach to achieving some measure of human good, without succumbing either to irrational invocations of the supernatural or to self-destructive scepticism, but with grace, dignity, and realism. Although I am not generally an optimist, I think Aristotle is right about moral progress. On the other hand, if the ecologists are right, population growth may sunder human progress in all its manifestations, moral progress included.

Acknowledgements. Thanks especially to Paul Churchland, Anne Churchland, Sarah Bellam, William Peabody, Steve Quartz, and David Sharp. This work was supported by a UC President's Humanities Fellowship, 1990–91.

REFERENCES

- Bjorklund, A., O.G. Nilsson, and P. Kalen (1990). Reafferentation of the subcortically denervated hippocampus as a model for transplant-induced functional recovery in the CNS. *Progress in Brain Research* 83: 411–426.
- Caudill, Maureen and Charles Butler (1990). *Naturally Intelligent Systems*. Cambridge, MA: MIT Press.
-

- Churchland, Patricia Smith (1986). *Neurophilosophy: Towards a Unified Science of the Mind-Brain*. Cambridge, MA: MIT Press.
- Churchland, Patricia Smith and Terrence J. Sejnowski (1991). *The Net Effect: Models and Methods on the Frontiers of Computational Neuroscience*. Cambridge, MA: MIT Press.
- Churchland, Paul M. (1984). *Matter and Consciousness*. Cambridge, MA: MIT Press.
- Churchland, Paul M. (1989). *A Neurocomputational Perspective*. Cambridge, MA: MIT Press.
- Gardner, Martin (1988). *The New Age: Notes of a Fringe-Watcher*. Buffalo, NY: Prometheus Books.
- Hecht-Neilson, Robert. (1990). *Neurocomputing*. Reading, MA: Addison-Wesley.
- Keynes, G., ed. (1966). *Blake, Complete Writings, with Variant Readings*. Oxford: Oxford University Press.
- Luker, Kristin (1984). *Abortion and the Politics of Motherhood*. Berkeley and Los Angeles, CA: University of California Press.
- Ojemann, George. (1983). Brain organization for language from the perspective of electrical stimulation mapping. *Behavioral and Brain Sciences* 6: 189-230.
- Ojemann, George, J. Ojemann, E. Lettich, and M. Berger. (1989). Cortical language localization in left, dominant hemisphere. An electrical stimulation mapping investigation in 117 patients. *Journal of Neurosurgery* 71: 316-326.
- White, Andrew Dickson (1896). *History of the Warfare of Science with Theology in Christendom*, vols. 1 and 2. Reprinted (1978). Gloucester, MA: Peter Smith.