8 Explaining the data

When social scientists maintain that their explanations are value-neutral, they usually mean that they are value-free—free of value-judgments. Explanation 1 is value-free, but 2 is not.

1 The legislator voted against the tax cut because he believed the cut to be unjust.

2 The legislator voted against the tax cut because the cut is unjust.

While 2 includes a judgment of the justice of the tax cut, 1 does not; it reports a fact about the legislator’s beliefs. Most social scientists would say that 2 is value-neutral. But explanations can be value-free and partisan. They can be partisan in favoring a particular conception of the good or in opposing the possibility of any one of a number of different conceptions of the good.

For example, in his book A Theory of Justice, John Rawls describes the circumstances under which justice is a moral virtue—what he calls “the first virtue of social institutions.”1 The circumstances of justice obtain, he says, when mutually disinterested persons put forward conflicting claims for the distribution of social advantages under conditions of moderate scarcity. In the absence of these circumstances, there is no occasion for the virtue.2 Imagine, then, that a social scientist offers explanations of the behavior of the members of a social group according to which the members are not mutually disinterested or never put forward conflicting claims for the distribution of social advantages. That is, the explanations, if true, show that the circumstances of justice do not obtain. Though the explanation says nothing about justice, it is not neutral on whether some distribution of advantages to members is just; for the explanation is incompatible with every judgment of the justice or injustice of a distribution of advantages to the members of the group. The explanation is neutral between whether the tax cut is just or unjust, but partisan on the question of whether the cut is a matter of justice at all.
To take another example, Karl Marx did not believe that justice was a moral virtue. According to his theory of history, concepts of right and justice are part of the ideological superstructure of society. He showed how they contribute to the mode of production of capitalism and how they rely on false assumptions about the freedom of labor in a capitalist economy. His explanations of the coercion built into the capitalist relations of production do not contain any judgments of the justice or injustice of capitalism; but they are, nevertheless, not neutral on questions of justice — for example, whether justice is a social virtue — for they oppose the assumptions — for example, those concerning the freedom of wage laborers — on which the judgments of justice and injustice depend.

My aim in this chapter is to show that many of the explanations offered by social scientists are partisan, not because they favor a particular conception of the good but because they are incompatible with a view of the subject on which many moral judgments and conceptions of the good depend. In particular, I argue that the explanations that many social scientists offer of the behavior of their subjects imply that our familiar moral concepts of responsibility, praise, blame, duty, obligation, respect, and resentment do not apply to them at all. If the social scientists' explanations of a subject's behavior are true, then many of our moral assessments of the subject or her behavior cannot be true, for the explanations imply that she lacks the kind of agency that many of our familiar moral concepts assume. First, however, I survey some of the prominent forms of explanation in the social sciences — namely, functional, covering-law, decision-theoretic, and hermeneutic — and explain how they oppose the rational explanations that the subject offers of her own behavior. Finally, I show how, in opposing her own explanations, these scientific explanations put the agency of the subject in doubt and, as a result, oppose most moral assessments of her conduct.

Scientific explanation

There is an idea common to a number of different social sciences that the participants in a social practice are benighted and that the accounts they offer of their own attitudes and actions are poor and should be discounted and replaced with other, richer, more enlightened or profitable lines of explanation. According to this idea, little or no credit should be extended to a participant's account of an action, and the entire inventory of these accounts should be marked down or written off as false or mistaken. Richard Nisbett and Timothy Wilson, for example, write in a paper in Psychological Review: “The evidence indicates it may be quite misleading for social scientists to ask their subjects about the influences on their evaluations, choices or behavior. The relevant research indicates that such reports, as well as predictions, may have little value except for whatever utility they have in the study of verbal explanations per se.”

This idea contrasts with our more ordinary, unscientific idea that while some participant accounts should arouse our suspicion, others should be afforded a full measure of our trust. According to our ordinary idea, we lose confidence in one account by gaining confidence in another, and when we set some accounts aside as false, we assume that others are the real coin. Participant accounts, according to our ordinary idea, should be marked down or depreciated not as a corporate body but singly, one account at a time.

The idea that participants do not understand themselves or that they are understood only by others invested with a science or a scientific theory opposes our ordinary practice of trusting and saving some participant explanations. It also opposes, I believe, our ordinary practice of treating participants as moral agents, as individuals who have moral rights and moral obligations or a conception of the good and towards whom it is appropriate to adopt a range of moral attitudes and emotions.

If the social sciences discount participant accounts, what is it that they are discounting? How do participants explain their own behavior? One common way is by offering a rational explanation. People usually explain what they do by citing their reasons for doing it. A rational explanation of a person's behavior explains the behavior if it cites an intention that is the cause of the behavior and that, under some description of the behavior, rationalizes it or makes it seem reasonable. The reasons or intentions that people cite usually include beliefs and desires that are meant to show what they did was a reasonable thing to have done in the circumstances.

Of course, the fact that people explain some of their behavior by saying what their reasons were does not mean that their reasons explain their behavior. Sometimes we have a good reason to discount or look beyond a participant's own explanation and replace it with a different one of our own. Discounting is part of our ordinary, and not merely our scientific, practice. However, the everyday reasons we have for discounting are reasons to discount some participant accounts but not all; for when, as part of our ordinary practice, we discredit one account, it is usually because we credit another. In fact, our ordinary practice of setting aside or discounting participant accounts is based on the assumption that false or valueless accounts are the exception rather than the rule; for we assume that the participants in social, political, and economic life are rational (even if their behavior is sometimes irrational), and to assume
that participants are rational is to assume that as a rule their own reasons for their behavior explain it.

We may discount a participant’s account because we believe that it is insincere and that she is hiding rather than citing the causes of her behavior; but part of our reason for doubting her sincerity is some prior knowledge of her intentions. Moreover, we may discount a participant’s account because we believe she is self-deceived and is hiding from herself the cause of her behavior; but, again, among our reasons for believing that she is self-deceived are our own beliefs about her intentions. However, what we know of people’s intentions we know, in part at least, because we believe that many of their accounts of their own behavior are true. When, as part of our ordinary practice, we replace a participant’s own rational explanation with an explanation of our own, we assume that normally her judgments about her own intentions are true. The more error we impute to her, the less we understand of her thought or her talk and the less reason we have to think that her intentional explanations of her behavior are mistaken. That is, our own understanding of what she says or thinks relies on the principle of charity, the principle that another’s judgments and intentions are not radically different from our own. According to the principle of charity, there is a limit to the mistakes that we can attribute to others in their judgments of their own intentions. The greater the difference between our judgments and theirs, the more reason we have to believe that it is we, rather than they, who are mistaken.

Many social scientists put aside or look beyond participant accounts, because they believe that these accounts do not explain enough. For example, some social scientists are interested in explaining not the behavior of individuals but their intentions or in explaining collective or large-scale social events or some statistical findings about a group. In these cases, when the social scientists put participant accounts aside, they do not discount the accounts in the sense of “discount” that I am talking about, for they do not say that the participants’ accounts are mistaken but only that they are not very deep or to the point. In such cases, the social scientists’ account and the participants’ accounts are compatible, for there is no implication that what the participants say is false. When I speak of discounting a participant’s explanation, I mean offering an explanation that is inconsistent with, not merely different from, the participant’s own. To discount is to imply that the reasons that participants offer in trying to explain their own behavior are not a reason for a cause of their behavior. For social scientists to discount, in this sense, implies more than that they know something the participants don’t; it implies that the participants have false beliefs about their own intentions.

When psychologists maintain, for example, that the causes of a subject’s behavior are not her intentions but processes or events that are hidden from her view, they discount a subject’s account of her behavior in the sense I am talking about. Thus, Latane and Darley, in their study *The Unresponsive Bystander: Why doesn’t he Help?*, discount when they maintain that bystanders who come or fail to come to the aid of people in distress are mistaken about the causes of their behavior. It is the presence or absence of other bystanders, according to Latane and Darley, rather than the intentions the bystanders cite, that explain their behavior. Storms and Nisbett, in their study “Insomnia and the Attribution Process,” discount when they argue that it is the presence of hidden causes – for example, an overheated room or a tendency to work or smoke before going to bed – that explains the behavior of insomniacs rather than the worries or concerns that insomniacs give as reasons for their not sleeping.

Social scientists who offer neurophysiological explanations of behavior for which subjects offer rational accounts, on the other hand, do not discount if the intentions that the subjects cite are type- or token-identical to the neurophysiological conditions or events cited by the scientists. Though the explanations are different, they are not inconsistent; for if an event in my brain is the cause of my arm going up and the brain event is identical with an intention of mine, then the intention causes my arm to go up as well.

Similarly, when Marx and Engels offer functionalist explanations of the behavior of factory-owners in England in the nineteenth century, they do not discount the owners’ own rational explanations of their actions; for although the owners might not accept the teleological explanations that Marx and Engels offer, the idea that their actions have a function or that the function is to increase the forces of production is consistent with the owners’ understanding that their intentions – for example, to increase profits – are the cause of their behavior. According to Marx and Engels, the factory-owners’ explanations are marked down not because what they say is mistaken but because they say too little: they do not explain the intentions.

Four models of explanation

Though most social scientists agree that the subject’s explanations are mistaken, their standards of correct explanation differ, and each standard offers a different reason for discounting the subject’s account. Emile Durkheim, writing on social causation, A. R. Radcliffe-Brown on ritual action, Sigmund Freud on the psychopathology of everyday life, and Gary Becker on the economic approach to human behavior each adopt a different standard – namely, covering-law (Durkheim), functional (Radcliffe-
Brown), hermeneutic (Freud), and decision-theoretic (Becker) — but the explanations of behavior that each offers are incompatible with the rational explanations that are offered by their subjects.

Durkheim stands in the Galilean tradition in the social sciences. It is his view that every event, whether in the natural or the social world, is to be understood in terms of its efficient cause. To explain why some individuals join a church, commit suicide, or remain celibate throughout their lives, one needs to show that acting so is connected in a lawful way to some social fact. The task of social science for Durkheim is to discover the causal laws that describe the links between social facts and individual behavior.

A. R. Radcliffe-Brown, on the other hand, stands in the Aristotelian tradition. It is his view that in order to explain why an individual engages in a religious practice or holds to a system of religious belief, it is necessary to discover the purpose or ultimate cause behind the practice or belief. The ultimate, rather than the proximate or efficient, cause is what explains the individual's behavior, and the ultimate cause is not the purpose of an individual but of a group.

Gary Becker also stands in the Aristotelian tradition. However, the purposes that explain behavior, on his view, are always individual purposes — for example, the preference of the consumer or producer for more of some good over less. According to Becker, the task of social science is to show that in most areas of their lives people behave as if they were economic agents and were attempting to maximize individual utility or profit, even if increasing utility or profit is not their intention.

Freud stands with one foot in each of these traditions. He believes that an individual's behavior is to be explained by its purpose, but that the purpose is an intended, though unconscious, purpose; and, according to Freud, an individual's unconscious purpose in doing something explains what she did only if it is the efficient cause of her doing it.

What Durkheim, Freud, Becker, and Radcliffe-Brown have in common is that each offers explanations of social life that are incompatible with all or many of the explanations offered by the participants in that life. According to each, participants, unless enlightened by a social theory or tutored by those who have mastered it, do not understand the facts of their life clearly or deeply enough to be able to explain large parts of that life to themselves or to others.

Durkheim's idea is that desires cannot be the cause of actions unless there is a lawlike relation between desires and actions. Since, whatever the desire with which an agent performs an action, there are other agents who share the desire but do not perform the action, this shows, on Durkheim's view, that desires are not the causes of action.

Durkheim's argument in *Suicide* is meant to be general. Only if the intentions that agents cite when they attempt to explain their behavior are always followed by the behavior are they the cause of it. Since the behavior does not always follow the intentions, the latter cannot be the cause of the former. And if the intentions are not the cause, then they do not explain the behavior.

Durkheim's reasoning proceeds from a principle of causality similar to that which Donald Davidson calls "The Principle of the Nomological Character of Causality": that if one event is a cause of another, then there must be a law relating them. If there are no laws relating an intention to an action, the intention cannot be the cause of the action; and if the intention is not the cause, then it cannot explain the action.

Durkheim replaced the language of reasons, the language that participants use to explain their own actions, with a language of causes that are not reasons. Freud, by contrast, used the language of reasons in his explanations of behavior; however, the reasons he cites are not the participant's own; for although they belong to the participant and are part of her intention, she is not conscious of those intentions.
According to Freud, a person has desires of a sexual nature that are incompatible with her thoughts of herself (with her ego). They arise from her animal instincts (her libido), and they conflict with the ethical or aesthetic standards by which she judges herself and her own worth (the superego); and, because of the conflict, the thought that she has these forbidden desires is too painful for her to bear. She cannot let on to herself, let alone to others, that such desires are her own. However, she cannot turn her back on them entirely, because these desires cannot forever remain unfulfilled.\footnote{33}

There are two mechanisms, according to Freud, by which a person can hide forbidden desires from herself and, at the same time, to some degree, satisfy them: repression and sublimation. Each defends the ego against the threat posed by the recognition of the desires, and each, at the same time, offers the desires an outlet. Through the mechanisms of repression and sublimation, the agent's forbidden desires are satisfied, though in a disguised and unrecognizable form.

The mechanism of repression, according to Freud, yields two kinds of actions: "faulty actions" and "inadvertent actions." Under the first label Freud includes acts of forgetting, slips of the pen or tongue, the losing or breaking of objects, bunglings, and misreadings. In doing these things, the agent is not conscious of having acted intentionally at all. For example, a young boy falls and, in falling, knocks over his parents' wedding picture. He sees this as an accident, as a movement of his body that was in no way intended. However, the boy, like any young boy, desires to replace his father in the conjugal bed (oedipal desire). This is a forbidden desire, and the boy has repressed it. Moreover, there is no description that the boy would apply to his behavior whereby the oedipal desire would give him a reason to fall and knock over the picture. The description under which the movements of his body are rationalized by that desire is a description that the boy would not apply to these movements: namely, "replacing father in mother's bed." The boy's oedipal desires are both a cause and a reason for his falling and knocking over the picture. However, the falling and knocking over are not understood by the boy to be an oedipal action, and, consequently, the boy is able to hide from himself that he has oedipal desires that cause and rationalize his movements.

Inadvertent actions, the second means of expressing repressed desires, are actions that the agents believe to be intentional movements but done without any aim or purpose. However, on Freud's view, they are not aimless, and the agents perform them with an intention, though not a conscious one, of satisfying a sexual desire. Freud cites humming tunes, fiddling with things, and fingering one's clothing as examples of inadvertent actions. When they finger or fiddle, agents don't believe that they are satisfying any sexual desires. However, the fingering or fiddling satisfies an unconscious sexual desire, although the description under which the behavior satisfies the desire is not a description that the agents believe to be true of their fingering, for it is not in any literal sense true of these movements of their body.

For example, a priest fiddles with his buttons while he listens to a man confess his sexual sins. He believes that he is fiddling with his buttons and that he is doing it intentionally, but he believes that he is doing it aimlessly and for no reason or purpose. However, the priest has a forbidden desire to do what the man confesses to having done, and, in fingering his buttons, the priest is satisfying this desire without seeming to do so. In fingering his buttons, the priest is removing his frock, the emblem of priestly privileges but also of priestly duties. These duties include abstinence. To be defrocked is to be stripped of this duty and to become, once again, a bare man who, at times at least, can satisfy his sexual desires. Understood this way, his movements are a means to a desired end. However, he doesn't understand them in this way and so does not believe that they are a means to any end that he desires. Through repression, the priest is able to satisfy his sexual desires and, at the same time, feel pure and free from sin. The priest does not literally remove his frock or engage in any sexual adventure, but, given the connection between the movement of his fingers on the buttons and the removal and the adventure, he is able to satisfy the desire for sexual adventure by remaining in his study or confessional and fingering alone.

Sublimation, according to Freud, is the other mechanism that allows agents simultaneously to hide and to act on their forbidden desires. With sublimation, the agents do not repress their forbidden desires but substitute for them desires that are culturally acceptable. Agents can act on these "acceptable" desires without revealing to themselves that they are acting on a forbidden desire. These acceptable desires are not just any stand-ins or substitutes for the forbidden ones; they are symbolic substitutes. Consequently, when the agents satisfy the stand-in, they satisfy the desire. That is, satisfying the stand-in is a means of satisfying the forbidden and hidden desire, on Freud's view, in virtue of the fact that the one is a symbol of the other.\footnote{44}

In his study of Leonardo da Vinci, for example, Freud explains how Leonardo sublimated his sexual desires into an urge for scientific research and, in particular, research on the nature of flight.\footnote{55} For this he chose to study birds, and his scientific papers include a detailed study of vultures.
Leonardo explained his studying vultures by citing his desire to learn about the nature of flight. However, according to Freud, this desire was a symbolic substitute for a forbidden, homosexual desire, and it is this forbidden desire that explains his interest in and study of these birds. On Freud's view, it is Leonardo's hidden or latent desire and not his manifest or cited desire, the desire he cites as his reason for acting thus, that is the cause of his behavior. Freud's explanation and Leonardo's are incompatible, and, as a result, Freud can be said to discount Leonardo's own explanations of his behavior.

In short, Freud discounts the accounts of participants because, on his view, many of their actions are faulty, inadvertent, or acts of sublimation. In the case of faulty actions, Freud discounts the participant's account because the participant believes that the behavior is unintentional. The movements of her body, so far as the participant understands them, are accidents; but according to Freud, she is mistaken. In the case of inadvertent actions, the participant believes that her movements are intentional, but she does not believe that she performs them for any reason or with any purpose; but here again the participant is mistaken. With both faulty and inadvertent actions, participants maintain that their actions were purposeless, while Freud discounts their view, maintaining that they were purposeful.

In the case of acts of sublimation, participants believe that their actions are purposeful; but Freud maintains that they do not understand the purposes that explain them. Only if Leonardo were to learn the psychoanalytic meaning of vultures would he be able to understand the purposes underlying his study. However, to understand this, Leonardo would need the theoretical language of psychoanalysis. According to Freud's theory of sublimation, there are wide expanses in our lives—namely, our creative or artistic efforts—which can only be rationally explained under descriptions that, without the aid of the theory, will seem to be false rather than true.

Radcliffe-Brown

According to Freud, the purposes that explain a participant's actions are often intended (though unrecognized) purposes. On the other hand, according to Radcliffe-Brown, the purposes are often unintended. In his studies of ritual behavior, Radcliffe-Brown sets aside the rational explanations that participants offer of their actions and focuses instead on the effects of the actions on the group. These effects, according to him, can be seen by the anthropologist but not by participants.

Let us suppose that we wish to investigate in Australian tribes the totemic rites of a kind widely distributed over a large part of the continent. The ostensible purpose of these rites, as stated by the agents themselves, is to renew or maintain some part of nature, such as a species of animal or plant, or rain, or hot or cold weather. With reference to this purpose we have to say that from our point of view the agents are mistaken, that the rites do not actually do what they are believed to do.26

My own view is that the negative and positive rites of savages exist and persist because they are part of the mechanism by which an orderly society maintains itself in existence, serving as they do to establish certain fundamental social values. The beliefs by which the rites themselves are justified and given some sort of consistency are the rationalizations of symbolic actions and of the sentiments associated with them.27

According to Radcliffe-Brown, in the case of ritual actions, it is always a mistake to explain the action by citing the participant's intentions.

The very common tendency to look for the explanation of ritual actions in their purpose is the result of a false assimilation of them to what may be called technical acts. In any technical activity an adequate statement of the purpose of any particular act or series of acts constitutes by itself a sufficient explanation. But ritual acts differ from technical acts in having in all instances some expressive or symbolic element in them.28

Participants in a rain dance, for example, explain their actions by citing their desire for rain. Early anthropologists like Sir James Fraser accepted these explanations at face value. They assumed that the dancing was a technical act and that the dancers desired rain and believed the dancing to be a cause of it. Finally, they assumed that the dancers' belief and desire were the cause of the dancing. Radcliffe-Brown objects. The dancing, he argues, is a ritual act, and it is unreasonable to think that the dancers have false beliefs about the causes of rain and that their false beliefs are the cause of their dancing.29 The dancing is to be explained, on Radcliffe-Brown's view, by an ultimate rather than by any proximate cause.

On Radcliffe-Brown's view, a participant's stated beliefs and desires do not explain rituals, while on Freud's view they do not explain art. However, while for Freud art is explained by hidden desires, for Radcliffe-Brown ritual is not explained by any desires at all.30 Radcliffe-Brown
believes that it is wrong to look for a rational explanation of a participant’s ritual actions, and because participants offer rational explanations, he discounts them.

**Gary Becker**

According to the economic approach to human behavior, people act so as to maximize expected utility on the basis of more or less stable preferences and coordinate their actions with the actions of others through markets. According to Becker, the economic approach is applicable not only to behavior within the monetary market sector but to family life as well. On his view, it is the one approach in the social sciences that provides a unified framework for understanding all human behavior.

All human behavior can be viewed as involving participants who maximize their utility from a stable set of preferences and accumulate an optimal amount of information and other inputs in a variety of markets.

I have come to the position that the economic approach is a comprehensive one that is applicable to all human behavior, be it behavior involving money prices or imputed shadow prices, repeated or infrequent decisions, large or minor decisions, emotional or mechanical ends, rich or poor persons, men or women, adults or children, brilliant or stupid persons, patients or therapists, business men or politicians, teachers or students.

Whenever people choose between doing A and doing B – for example, marrying or staying single, working or stealing for a living, having a child or remaining childless, joining or not joining a church, becoming a Democrat or a Republican – they will choose A over B, according to the economic approach, if and only if the expected utility of A is equal to or greater than that of B. So, for example, people will marry if they can expect to raise their utility level above what it would be were they to remain single; and should the supply of mates increase, the cost of a mate will decrease, and, other things being equal, the number of marriages will rise.

Becker does not assume that participants intend to maximize utility or can say what the expected value of one choice is over another. It is enough that they act as if they calculated and compared these values; it is not necessary that they understand themselves to be acting in this way. Most people do not view their decision to marry, steal bread, or bear children the way that Gary Becker does. However, according to Becker, this does not mean that their “personal” decisions are not a function of costs and benefits: “The economic approach does not assume that decision units are necessarily conscious of their efforts to maximize or can verbalize or otherwise describe in an informative way the reasons for the systematic patterns in their behavior.”

Though the explanations of family life that Becker offers are different from those that are offered by most participants, it is not clear whether Becker, in offering his economic explanations, is discounting the accounts of the participants. That is, it is not clear whether what Becker says about their behavior is true, the participants’ own rational explanations must be mistaken. It is not clear, because Becker sometimes speaks as if the economic approach describes the causes of a subject’s action and, at other times, as if it merely describes the action.

Subjects can be economic in either of two senses: one causal, the other not. They are economic in a noncausal sense if their choices are consistent with the maximization of a well-ordered function such as a utility or expected utility function, and in a causal sense if their choices are consistent with the maximization of the function because they intend them to be. A business firm, for example, might decrease its production in a way that increases profit, even if what causes it to act in this way is not an intention to profit; for increasing profit might be an unintended consequence of the firm’s behavior, and an economist might explain the firm’s choice by citing the effects on profit and ignore altogether the questions of intention. In this case, though the economist ignores the firm’s intentions, she is not discounting the firm’s own, intentional account of its choice, for it is possible for both her account and the firm’s to be true. Only if the economist explains that an intention to increase profits was the cause of the cut in production, while the firm explains that it was not, does her explanation discount the firm’s, for only then are the two explanations incompatible.

However, this view of the economic approach is not entirely adequate, for there are two reasons to think that the question of whether subjects act so as to maximize some well-ordered function cannot be separated from the causal question of how they choose their action. First, effects on profits explain choice only if the relation between choice and profit is lawlike; for only then is the effect on profit a reason to expect the choice. But an account of the causes can be a reason to believe that the relation between choice and profit is not lawlike; for if, on the firm’s own account, its choices are caused by intentions that are known not to result, as a rule, in any increase in profit, then there is reason to believe that the relation between choice and profit is not lawlike.
I offer four arguments for this conclusion. The first, the argument from responsibility, is the most general and abstract. Here I argue that our concept of moral responsibility is such that it is only to subjects whose behavior is caused to a high degree by their conscious intentions that we can attribute moral responsibility. My argument here resembles the soft determinist's argument that moral responsibility is not only compatible with causal determinism but inconceivable without it. The second, the argument from conscious interests, and third, the argument from the good will, are less general. The point of these arguments is that the subject in the social sciences cannot have the interests assumed by Mill's utilitarianism or the will assumed by Kant's account of moral worth. Finally, in the fourth, the argument from moral emotion, I draw on an idea of Peter Strawson's that, as part of our moral practice, we adopt participant attitudes and feelings toward others, and argue that we cannot adopt these attitudes toward others whose conscious intentions do not, as a rule, explain their behavior.

The argument from responsibility

To be a moral subject, much of a subject's behavior must be caused by her conscious intentions, for moral subjects are bearers of moral responsibility; but a subject bears moral responsibility for her behavior only if she is consciously rational, and she is consciously rational only if, as a rule, her conscious intentions are the cause of her behavior.

Some moral philosophers have argued that a person is morally responsible for what she does only if she would have done otherwise had she so intended. I am claiming less. According to my argument, a subject is morally responsible for doing A only if some large (perhaps weighted) number of the things she does (which need not include A) would have been otherwise had her conscious intentions been otherwise. On this view, we can hold someone morally responsible for an unintended or involuntary movement – for example, falling down or falling asleep – as long as enough of her other movements are intended or voluntary. The more a subject's movements are unresponsive or not subject to her conscious intentions, the less appropriate it is to hold her morally responsible for any one of her bodily movements. The reason for this, I contend, is that unless a large enough number of her movements are understood to be caused by her conscious intentions, we lose our reason to believe that she is a consciously rational subject; and unless we have reason to believe that she is a consciously rational subject, we have no reason to attribute intentions to her at all. Subjects to whom we do not have good

The subjects of morality

When we try to explain the behavior of people, we assume that, as a rule, they are able to explain it to themselves. Though we judge that some of their explanations are mistaken, our judgment is based on the belief that others are correct; for there is a limit to how much error we can attribute to a person's understanding of her own actions. Many social scientists, however, assume that the explanations their subjects offer are always or often mistaken. Yet, our practice of making moral judgments rests on our ordinary standards of explanation rather than on the social scientists'. As a result, as I argue in the following paragraphs, the subjects of morality cannot be as often mistaken about the causes of their own behavior as these social scientists assume, and, as a result, the subjects of morality cannot be the subjects of the social sciences.

How frequently must a subject's conscious intentions explain her behavior for her to be a subject of morality? There is no sharp line here, no set number of mistakes that separates the subjects of the social sciences from those of morality. At one extreme is the thoroughly benighted subject: one whose conscious intentions are never a cause of her behavior. Durkheim's subjects are closer to the extreme than Radcliffe-Brown's. At the other extreme is the thoroughly lucid subject: one whose conscious intentions are always the cause of her behavior. The closer the moral subject is to the one extreme and the social science subject to the other, the clearer it is that they are incompatible. My position is that Durkheim's, Freud's, Becker's, and Radcliffe-Brown's subjects are close enough to one extreme and often benighted enough to make them strangers to morality.

Second, utility, unlike profit, is a psychological magnitude; it is a measure of preference, and preference is a state of mind that is identified by the participant's own account of her actions. If we accept a participant's account that she is choosing a less over a more preferred alternative, we must conclude that she is not acting to maximize utility. She is doing what she intends to do but not what she prefers to do. However, the economic approach to human behavior assumes that participants never choose less over more and so only intend to do what they prefer to do. Consequently, when the participant explains that her aim in choosing B over A is less over more and Becker explains her choice economically, he attributes preferences to her that are inconsistent with those that she attributes to herself. As a result, his economic explanation discounts her own account, even though it appears to be silent on the causes of her choices.
reason to attribute intentions are not the bearers of moral responsibility and, in that respect, are not subjects of morality.

Durkheim's subjects are not consciously rational, for were their intentions different, their behavior would remain the same. In arguing for social over psychological causes of behavior, Durkheim denies that there is a causal connection between conduct and intention and, as a result, offers us a subject whose behavior is not causally responsive to changes in her intentions. Such a subject is not a moral subject, and so Durkheim's subjects are not the subjects of morality.

Freud's subjects, unlike Durkheim's, are moved by psychological rather than social causes. Freud, unlike Durkheim, does not deny that intentions are explanatory; what he denies is that the intentions that participants cite are as a rule explanatory. According to Freud, there is no break or gap between intention and conduct, but only between conscious intention and conduct.

Because Freud's subjects, unlike Durkheim's, would very often do otherwise were they to intend otherwise, they would seem to be the subjects of morality. However, they are not moral subjects, for the intentions to which their behavior is responsive are too often intentions of which they themselves are not conscious. That is, the descriptions under which the behavior of Freud's subjects is intentional are very often descriptions under which they do not consciously intend their behavior. They very frequently hide from themselves the intentions that are the practical reasons for, and the causes of, their behavior. Moral responsibility, however, requires conscious rationality. It is only if some large (perhaps weighted) amount of a person's behavior varies with her conscious intentions that we hold her morally responsible for her behavior.

My argument is as follows. According to our idea of moral responsibility, morally responsible subjects are able to benefit from moral instruction. Given what they regard as very good reasons for revising their intentions, they can be expected to revise their behavior. The aim of moral instruction is to change what people do by changing their reasons for doing it. Freud's subjects are not the subjects of morality, because although their conscious reasons are open to moral instruction, these are very often not the cause of their behavior, and although their unconscious reasons are often the cause, these are not open to moral instruction, for the ego's mechanism of defense keeps the subjects from acknowledging that these unconscious reasons are theirs. "An unconscious wish cannot be influenced and it is independent of any contrary tendencies, whereas a conscious one is inhibited by whatever else is conscious and opposed to it." When moral instruction causes changes in the conscious intentions of Freud's subjects, it often does not cause changes in their behavior, for the behavior is not caused by these intentions; and because the ego employs evasive strategies that screen the unconscious causes from the subjects' consciousness, it screens them from moral instruction.

It is a long supberceded idea, and one derived from superficial appearance, that the patient suffers from a sort of ignorance, and that if one removes this ignorance by giving him information (about the causal connection of his illness with his life, about his experiences in childhood, and so on) he is bound to recover. The pathological factor is not his ignorance in itself, but the root of this ignorance in his inner resistances; it was they that first called this ignorance into being, and they still maintain it now. The task of the treatment lies in combating these resistances. Informsing the patient of what he does not know because he has repressed it is only one of the necessary preliminaries to the treatment. If knowledge about the unconscious were as important for the patient as people inexperienced in psychoanalysis imagine, listening to lectures or reading books would be enough to cure him. Such measures, however, have as much influence on the symptoms of nervous illness as a distribution of menu-cards in a time of famine has upon hunger. The analogy goes even further than its immediate application; for informing the patient of his unconscious regularly results in an intensification of the conflict in him and an exacerbation of his troubles.47

In short, the subjects of Freud's double consciousness are not the subjects of morality, because the doubling and the inner resistances hide the causes of their conduct and prevent them from becoming objects of reason or moral reform.

Leonardo, for example, at least in his artistic life, is not a moral subject; for, on Freud's view, he cannot understand that his artistic endeavors are acts of sublimation or that the homosexual intentions causing them are his own intentions. He cannot understand, because the intentions are hidden in his unconscious and because repression and sublimation include a mechanism - namely, resistance - that causes him to deny that a forbidden intention is his.48

Becker's subjects resemble Durkheim's. We have no reason to believe that were their intentions different, their behavior would be different; for, on Becker's view, a person will act so as to maximize utility even if he does not intend to do so and even if he intends not to. The subjects of economics can act with altruistic intentions, but the actual, as against the intended, effect of their action is to increase their own utility. Persuading Becker's subjects to act with altruistic rather than egoistic intentions will
not cause them to act other than economically or to act in a way that does not maximize utility.

Radcliffe-Brown's subjects resemble Freud's. Like Freud's, they lead a double life. When they engage in technical acts, they are consciously rational subjects. When they engage in ritual actions, they are not. So, as participants in the rituals of life, they are not moral subjects, for in being subject to ritual, their actions are not subject to their own intentions. Like Freud's subjects of repression and sublimation, the subjects of ritual actions are not able to benefit from moral instruction. Given what they regard as very good reasons for revising their intentions, they cannot be expected to revise their behavior, for their behavior is not caused by their intentions but, in some mysterious fashion, by the effects of the behavior on the structure of their group.

Moreover, given that the judgments by subjects of ritual about their own intentions are so often mistaken, there is reason to doubt that they are rational subjects or that we understand their intentions. Radcliffe-Brown objects to Sir James Fraser's interpretations of the participants' intentions, because they are uncharitable - Fraser attributes too many mistaken beliefs about nature to the dancers. But in discounting the dancers' accounts, Radcliffe-Brown is being uncharitable as well, for he is attributing too many mistaken beliefs to the dancers about the cause of their own behavior. A subject who, in some large and central area of her life, is not rational or whose thought and behavior in that area are not interpretable is not (or at least in that area is not) a subject of morality.

The argument from conscious interests

There are some moral theories according to which what makes an action right is that it satisfies human desires. This was Mill's view: an action is right insofar as it adds to the satisfaction of human desires and wrong insofar as it detracts from them.49 Mill assumed that the satisfaction of more human desires makes the world a better place than the satisfaction of fewer.

Mill says in On the Logic of the Moral Sciences that states of mind - that is, thoughts, emotions, and actions - are caused by other states of mind or by states of body.50 He, like Durkheim, accepts the principle of the nomological character of causes. He believes that if one state of mind causes another, then the one must be regularly followed by the other, and that there must be a psychological law that links them. However, unlike Durkheim, he believes that there are such laws and that a science of psychology (he calls it a science of human nature) is possible; unlike Durkheim too, he accepts participant explanation, because he believes that the intentions that the participants cite are, as a rule, the causes of their behavior.

Like other states of mind, beliefs and desires, for Mill, are characterized by the role they play in causing behavior. According to Mill's philosophy of mind, states of mind that are never a cause of a person's thoughts, emotions, or actions are not states of belief or desire. So, on Mill's view, in denying that a person's intentions are a cause of her behavior, Durkheim is denying that she is in a state of mind that, according to utilitarianism, it is the business of morality to advance. No one is in a state of mind that, according to Mill, morally matters - namely, happiness - unless that state is one that causally matters.

Mill, like Davidson, believes that causality and intentionality go hand in hand.51 Part of our reason for attributing intentions to a subject is that they are a cause of her behavior. Belief and desire, on his view, cannot be epiphenomenal or nomologically inert. So, if Durkheim is correct in maintaining that social rather than psychological facts explain a subject's behavior, the subject would not be a subject of utilitarianism, for she would not have the kind of mind that makes states of happiness possible.52

There is another reason for thinking that Durkheim's subjects are not the subjects of Mill's utilitarianism: only those subjects whose actions are explained by their intentions can choose to conform their behavior to the prescriptions or advice of utilitarianism. That is, if a person's intentions are not a cause of her actions, then the intention to increase human happiness could never move her to action. In such a case, utilitarianism could not be a theory to guide human conduct. At best, it would be a theory by which to evaluate it; but Mill thought that utilitarianism was both.

Mill's utilitarianism is incompatible with Becker's social theory as well as Durkheim's. Though Becker and Mill both assume that their subjects have a coherent set of desires or preferences, Becker imputes preferences to his subjects that discount their own account of their behavior; for in offering his economic explanations, Becker attributes preferences to them that are not consistent with those that, in offering their account, they attribute to themselves.

There are two issues here. One is whether Becker can so often overrule his subjects' judgments about their own preferences. If, as Davidson argues, the bearers of intention must be largely correct in their judgments of intention, then he cannot; for the more he discounts his subjects' own judgments, the less he can attribute preferences to them at all.

The second issue is whether Becker's subjects can choose to conform
their behavior to the prescriptions of utilitarianism. Can a person who always acts to satisfy his preferences be caused by the belief that A is more conducive to human happiness than B to choose A? The answer to this question depends on the economist's concept of preference. As "preference" is traditionally defined in economics, each participant's preferences are independent of every other's. According to this definition of "preference," the subjects of economics are purely self-interested. They are not moved by feelings of sympathy or envy. Purely self-interested subjects can act as if they were utilitarians, but their reasons for acting cannot be utilitarian. That is, their actions cannot be caused by the (ultimate) desire to increase the happiness of others, for that is precluded by the economist's assumption of independence. If the subjects of utilitarianism must be able to choose human happiness for the sake of human happiness, then they cannot be the subjects of such an economic theory.

Radcliffe-Brown's subjects lead two lives, and in one, their ritualistic life, they are the bearers of human happiness; for in that life, their behavior is not subject to their beliefs and desires, and thus, in that life, there is reason to doubt that they have any desires at all. In addition, in their ritual lives, Radcliffe-Brown's subjects cannot act from any utilitarian desires, for their desires, if we can speak of them at all, are inert— are not a cause of any of their behavior.

The argument from the good will

Kant's moral theory, unlike Mill's, is not based on the effects of, but on the reasons for, an action. What gives an action moral worth, according to Kant, is that the agent performs it for the right reasons: out of a sense of duty. Moreover, duty may require a person to perform an action that is contrary to her own inclinations or desires; for, if it is morally right for a person to do something, it is right even if she would be happier (satisfy more of her desires) if she acted otherwise. Kant writes:

Only something which is conjoined with my will solely as a ground and never as an effect—something which does not serve my inclination but outweighs it or at least leaves it entirely out of account in my choice—and therefore only bare law for its own sake, can be an object of reverence and therewith a command. Now an action done from duty has to set aside altogether the influence of inclination, and along with inclination every object of the will; so there is nothing left able to determine the will except objectively the law and subjectively pure reverence for this practical law, and therefore the maxim of obeying this law even to the detriment of all my inclinations.33

It is part of Kant's moral psychology that a subject can be moved to act by the belief that she is morally obliged to and in the absence of a desire to; for, as Kant sees it, the absence of a desire is never a reason or an excuse for not doing what duty requires.34 On Kant's view, the moral quality of a person's action is dependent on the reasons why she performed it. Only if her reasons were moral reasons are her actions worthy of our moral respect. To comply with the moral law is a moral reason, according to Kant, while to increase one's own happiness or even the happiness of others is not. Consequently, if the intention to increase happiness, rather than the intention to comply with the law of promise-keeping, explains why a person keeps her promise, her will to keep the promise is not a good will. Her intention to do what the moral law requires must explain why she keeps the promise before her will to keep it can be said to be good.

If a participant's reasons do not explain her behavior, her reverence for the law does not explain it. Consequently, in discounting wholesale the participant's account and in maintaining that her reasons do not explain her behavior, the social scientist denies that the participant's will is a good will. In discounting wholesale the participant's account, the social scientist removes our ground for believing that the participant is a consciously rational subject and that she is capable of acting from any of her reasons, and, a fortiori, that she is capable of acting from her reverence for the law.

To deny, as Durkheim does, that reasons or intentions explain is "to deny to the concept of morality all truth and all relation to a possible object."35 The will of Durkheim's subject is never autonomous but always heteronomous (subject to external causes): it is never moved by the subject's intentions and thus never moved by her intention to comply with the moral law. As a result, the will of Durkheim's subject is never a good will, and his subjects are not the subjects of Kant's morality.

Freud's theory of double consciousness is also inconsistent with Kant's moral psychology; for if a subject's actions are often explained by desires that are buried in her unconscious and not by her conscious intentions, then she cannot act out of reverence for the law, for even if she has the conscious intention of doing what the law commands her to do, she is not acting with a good will, for it is the hidden desire rather than the intention that explains her action. Whenever Freud's subjects act from latent desires, their will is heteronomous. But because Freud's subjects so often act from latent desires, we lose our reasons for believing that...
their will is sometimes autonomous and that they sometimes act with the intention of complying with the law. We lose our reasons for believing that we understand their intentions or for believing that they sometimes act out of duty rather than to satisfy a sexual desire. As a result, there is reason to believe that Freud's subjects, no less than Durkheim's, are not the subjects of Kant's morality.

The point here, as in the preceding arguments, is that to accept a theory of human behavior according to which a subject's conscious intentions seldom explain her behavior requires a revision of our moral views. It requires that we revise that part of our moral view that is utilitarian and that we give up that part of our view that is Kantian as well. The psychology of Durkheim, Freud, Becker, and Radcliffe-Brown is not consistent with the moral theories of writers like Kant and Mill, for we cannot attribute to the subjects of their psychology the motivation needed for the behavior of these subjects to have moral worth or be conducive to human happiness. To discount, as a matter of general practice, participants' accounts of their own behavior is to deny morality, whether Kant's or Mill's, to its subjects.

The argument from moral emotion

Focused as it is on the notion of duty and on the idea of a moral law, Kant's moral philosophy emphasizes that part of our ordinary moral practice that is general and impersonal. It sets aside those emotions or sentiments which, in ordinary life, lead us to treat some subjects differently and more considerately than others. In his paper "Freedom and Resentment," Peter Strawson considers that part of our ordinary moral practice that takes some of these feelings and emotions into account. Strawson observes that although some of our ordinary moral practices require us to adopt a detached attitude toward others, the kind of attitude that Kant recommends, some other of our moral practices require us to adopt more personal and less detached, participant attitudes. Instead of viewing subjects disinterestedly and measuring their performances against an impersonal standard of duty, we act out of friendship or personal concern. Our less detached or more personal attitudes are what make our interpersonal relationships possible. Strawson includes as examples such attitudes or feelings as forgiveness, gratitude, resentment, love, good will, esteem, contempt, and anger.

According to Strawson, there are two kinds of consideration which serve to modify or temper our normal participant attitudes or feelings toward subjects or their actions. One serves to discount the action, the other to discount the subject. The first leads us to excuse subjects for a harm that they might have caused us, without inviting us to think that they are the kind of subject toward whom it would be inappropriate to ever feel moral emotions like anger, gratitude, or resentment. The fact that their actions were inadvertent, unavoidable, or done without malice, for example, is a reason not to resent them for causing us this harm, but it is not a reason to think that they are subjects whose actions should never give us a reason to be angry or resentful.

The second consideration, by contrast, invites us to view a subject as someone towards whom it is never appropriate to adopt any participant attitudes at all. To believe, for example, that a man is insane or thoroughly irrational is to have a reason to suspend our participant attitudes toward him altogether—a reason not to get angry with him if he behaves badly or to get indignant with him if he seems insensitive to our feelings or concerns. We may be harmed by his boasts, lapses, or lies, but if he lies or boasts compulsively, we should not resent him or feel contempt. Unless we believe that a man is rational and that his conscious intentions, as a rule, are the cause of his behavior, we should not make him an object of our moral attitudes and feelings.93

Strawson's interest is in the thesis of determinism. And the question he asks himself is whether a belief in that thesis would give us a reason to suspend our participant attitudes toward subjects whom we classify as rational.94 Would we have any reason to modify our ordinary practice of adopting participant attitudes toward some subjects and suspending them toward others if we believed that all actions and attitudes are causally determined? Strawson thinks that belief in such a general thesis would not give us any reason to withdraw these attitudes from every subject, rational and irrational, sane and insane alike. Belief in determinism, Strawson argues, is not a ground for switching from a participant to an objective attitude.

For it is not a consequence of any general thesis of determinism which might be true that nobody knows what he's doing or that everybody's behavior is unintelligible in terms of conscious purposes or that everybody lives in a world of delusion or that nobody has a moral sense, i.e., is susceptible of self-reactive attitudes, etc.95
preceding states of mind or events does not give us a reason for withholding feelings of affection, loyalty, or friendship towards her.

My concern, however, unlike Strawson's, is not whether determinism and resentment are compatible, but whether the social science explanation and resentment are, and my position is that they are not. Though it is not a consequence of any general thesis of determinism, it is a consequence of a wholesale discounting of participant accounts that nobody knows, to use Strawson's words, "what he is doing or that everybody's behavior is unintelligible in terms of conscious purposes or that everybody lives in a world of delusion or that nobody has a moral sense."

The consideration that the participants' own reasons as a rule do not explain their behavior (unlike the consideration that every thought and action is causally determined) is a reason, given our present moral practice, to suspend our participant attitudes. We cannot appropriately feel resentment or gratitude towards subjects who behave in a way that harms or benefits us but whose behavior often has little to do with their conscious intentions.

Because they so seldom understand the causes of their own behavior, Durkheim's, Freud's, and Becker's subjects are not the proper objects of anger, resentment, or respect; they are objects that we can observe, survey, examine, study, classify, motivate, supervise, regulate, command, discipline, treat, train, rehabilitate, or cure. A subject whose behavior is seldom explained by her own reasons for so acting is someone whom we might view as a burden, a ward, a client, a dependent, or a charge, but not as someone with a mind of her own, someone whose movements are actions of which she is the agent. We can try to improve her prospects by pulling her with carrots or pushing her with sticks, but not by putting stock in her own accounts of the causes of her own behavior.

If we interpret a subject out of our moral community by discounting her own explanations of her actions, we interpret her not only as someone towards whom we should not feel anger or resentment, but, in addition, as someone from whom we cannot expect anger or resentment either; for these emotions include intentions that can be imputed only to someone whose actions, as a rule, are caused by her intentions.

To discount a subject's own account of herself is to deny that she is a moral subject in two ways. It is to withhold from her our own moral emotions and also to withdraw from her her own emotions; for to discount her self-understandings is a reason for us not to feel resentment or anger towards her, but it is also a reason to think that she feels no resentment or anger towards us.

This last point is made by Marilyn Frye, who, in an essay entitled "A Note on Anger," points out that to be angry is to make a judgment. ⁴⁰

When I am angry with people for lying or boasting, I judge that they ought not to have lied or boasted. Frye's interest is in how we react to the anger of others, and, in particular, the ways we have of not taking their anger seriously. One way is to assume that the anger is not explained by their reasons for being angry, their judgment that someone did something wrong, but rather by some cause that is hidden from their view. ⁴¹

Frye discusses the role that gender plays in our practice of discounting participants' accounts of their emotions. In our culture, she observes, a woman is more likely than a man to have her own explanation of her anger (and her other moral emotions) discounted. A man who explains that the reason why he is angry is that he believes that he has been treated unfairly is more likely to have that accepted as an explanation of his anger than a woman is. In ordinary life, we are inclined to think that women, more than men, do not understand themselves.

If, in ordinary practice, we are less respectful of a woman's understanding of herself than we are of a man's, social theories like Durkheim's, Freud's, Becker's, and Radcliffe-Brown's change that. Men and women get equal treatment. The theories look upon men as men are inclined, in ordinary life, to look upon women. According to the theories, men have no more understanding of the causes of their own behavior than women do. Without the aid of theory or therapy, men and women stand together in the dark.

In turning aside the subject's own explanation of her anger and looking for a nonrational or hidden cause of it, we remove her anger and interpret her behavior as a symptom of something, rather than as a reasonable response to the wrongful actions of others. By turning her explanations aside, we retreat from the strains of involvement, for we hide her attitudes from our view; and, if we turn enough of her explanations aside, we will no longer see her as rational or as the bearer of any attitudes that deserve our respect.

It is not surprising, then, that people should feel demeaned or at risk when their own explanations of their actions are judged to be false and are replaced by explanations in which their reasons are not cited as the causes of their actions. For when many of a person's own explanations are discounted, her actions lose their moral worth, and she loses her status as a person capable of rationally deliberating among differing ends - as an autonomous individual - and a subject of our morality.

For some liberals - for example, J. S. Mill - value-neutrality is intended to protect and promote the autonomy of the individual. On their liberal view, education should produce in children the desire and the ability to rationally deliberate among differing ways of life and to make practical and moral decisions on their own; and the state should encourage
individual initiative and personal choice and resist all forms of paternalism. Society's basic institutions, they believe, should be designed to ensure that individuals, so far as possible, have the opportunity to express their nature as free and equal rational beings.

As I explained in chapter 1, the liberal sciences, like the state and the schools, are part of that design. Like other liberal institutions, they assume that deliberation among differing ends should be left to the individual to decide. However, as I have shown in this chapter, the standards of explanation common to the liberal sciences oppose the very idea of the autonomous individual. The standards take away what neutrality aims to protect: a community of free and equal rational beings legislating their own principles of conduct. In chapter 6, I explained how the standards for collecting data in the social sciences include devices for limiting the active presence of the subject in the practice of the science and, in particular, devices for weakening the will of the experimental or interview subject. The standards for explaining the behavior of the subject also include devices for limiting her active presence in the practice of the science. But whereas in collecting the data, the agency of the subject is limited or tightly controlled, in explaining the data, her agency is entirely eliminated, and she is no longer an individual with a conception of the good; but if the subjects of the social sciences have no conception of the good, their autonomy is no reason for the sciences to be neutral between the different conceptions of their subjects.

There is an incoherence in the ideals of the social sciences. On the one hand, the ideal of value-neutrality assumes that the subjects of the science are rational beings who choose between competing conceptions of the good and that, as rational beings, their choices are deserving of respect. On the other, the ideals of reliable and valid data and scientific explanation assume either that they are not rational beings or that their choices are not deserving of respect. What one ideal gives, the others take away. A coherent approach would either not include value-neutrality as a regulative ideal or not include the many devices for limiting the active presence of the subject in the social sciences.

Conclusion

What I have tried to show is that an approach to explanation common to a number of different social sciences is not consistent with a view of subjects assumed by many of our ordinary moral concepts. In particular, I have argued that if a subject's reasons are too seldom accepted as an explanation of her behavior, then we cannot adopt our ordinary moral attitudes toward her or view her as an appropriate bearer of moral responsibility. Whereas in other chapters, I have shown how particular practices in the social sciences favor one conception of the good over another, in this chapter, I have shown how the practice of explanation in the social sciences opposes the very idea that subjects have conceptions of the good that institutions, like the social sciences, should respect and neither condemn nor condone.

My claim was not that determinism and morality are incompatible, but that a particular form of misjudgment and morality are so. It is inconsistent to say, I argued, that someone is a moral subject yet mistaken in many of her judgments about the immediate causes of her own behavior. In short, the point of this chapter has been to defend a new version of incompatibilism – the thesis that discounting the social sciences is incompatible with our ordinary or everyday psychology of participant explanation and with the groundwork of our moral philosophy – and to show that although the social sciences are founded on the liberal ideal of value-neutrality, the approach to explanation common to the social sciences is illiberal, for it does not recognize – let alone respect – the conception of the good of the subjects of the science and, far from encouraging or endorsing their autonomy, opposes the idea that they have any autonomy at all.

NOTES

1 Rawls, Theory of Justice, p. 3.
2 Ibid., p. 128.
3 See S. Lukes, Marxism and Morality (Oxford University Press, 1985), for a discussion of Marx's criticisms of these concepts. In some of his criticisms, Marx seems to suggest that all moral concepts are bogus and simply mask the reality of the social relations of production under capitalism; but I am persuaded by Lukes that Marx does not mean to call all moral concepts into doubt but only the juridical concepts favored by the many socialists of his day.
5 R. E. Nisbett and T. D. Wilson, "Telling More than we can Know: Verbal Reports on Mental Processes," Psychological Review, 84 (1977), p. 247. Nisbett and Wilson argue that accounts offered by subjects of the causes of their own judgments, as well as the causes of their behavior, are very often mistaken. My remarks are limited to judgments that people make about the causes of their own behavior.
6 I want to allow that our ideas about how to explain our behavior and about when such explanations are in order have a history. There is no reason to believe that the explanations that people offer of their behavior are everywhere
the same, or even that explanations are everywhere offered. That is, the practice of calling on and being called on by others to explain one's behavior may be a feature of our own time and place. I do not know, e.g., that our practice of seeing human beings as agents and as acting on the basis of reasons was also the practice of the serfs of thirteenth-century Europe or is now the practice of the Trobriand Islanders. Moreover, though I will contrast our ordinary practice with the practice of social science, I want to admit that our ordinary practice is affected by the practice of social science. We are influenced by expert opinion and may come to think about ourselves the way the experts do. If the experts maintain that you and I are utility-maximizers, we may increasingly become so. This fact alone, as Foucault argues, opposes the idea that the social sciences are morally neutral and that a theory in the social sciences can be positive without, at the same time, being critical. My point is merely that, at present, in my own part of the world, there is a contrast between the ordinary, garden variety explanations that people offer of their own behavior and the explanations that the social sciences offer, and that the garden variety are rational explanations.

7 Donald Davidson, in Essays on Actions and Events (Oxford University Press, 1980), essays 1–5, offers a subtle account of rational explanation. He argues that reasons can be causes and that rational explanations are a species of singular causal explanation. According to this view, a participant's rational explanation is false or inadequate if the reasons she cites do not cause her behavior, or cause it in just the right way. Social scientists who believe that only causes explain and that reasons are not causes, will discount or set aside as false a rational explanation, whether offered by the participant or an observer, as I explain later in this chapter.

8 For a more detailed discussion of this point, see D. Davidson, Inquiries into Truth and Interpretation (Oxford University Press, 1984), essays 9–11, and Actions and Events, essay 15. Davidson argues that to interpret a subject's behavior and attribute thoughts to her, we must assume that her judgments about her own thoughts are, for the most part, true. Davidson writes: "What operates as a constraint on the interpreter amounts to a bestowal of authority on the person interpreted; when he honestly expresses his motives, beliefs, intentions, and desires an interpreter must, if he wants to understand, interpret the speech in a way that makes sense not only of what the speaker says but also of what he believes and wants and does. Making such sense requires that causality and rationality go hand in hand in important cases: perception, intentional action and the indirect passions are examples. People are in general right about the mental causes of their emotions, intentions, and actions because as interpreters we interpret them so as to make them so. We must, if we are to interpret them at all" (Actions and Events, p. 290). On Davidson's view, to discount many of a participant's accounts is to lose our reasons for believing her to be an agent and for believing that we understand the contents of her thoughts or intentions. See also my work "Davidson and Social Science," in Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson, ed. E. Lepore (Basil Blackwell, 1986), pp. 272–304. For an opposing view of the relation between rationality and charity, see P. Thagard and R. E. Nisbett, "Rationality and Charity," Philosophy of Science, 50 (1983), pp. 250–67.


11 For a discussion of the compatibility of Marx's materialist explanations and the rational explanations that individual workers and capitalists might offer of their own behavior, see Wood, Karl Marx, pp. 101–22.

12 This idea belongs to G. H. von Wright, Explanation and Understanding (Cornell University Press, 1971), pp. 1–33.


14 Radcliffe-Brown, Structure.


18 Durkheim, Suicide, pp. 297–8.

19 Ibid., p. 306.

20 Davidson, Inquiries, p. 208.

21 On Durkheim's view, beliefs and desires do not explain an action, but they can help to define it. Only those actions that subjects perform in the belief that they will bring about their own death, e.g., count as suicide, according to Durkheim. Though there is no causal connection between the belief and the action, on Durkheim's view, there is a definitional one. See Suicide, p. 44.

22 See, e.g., Freud's remark in The History of the Psychoanalytic Movement that he came to realize that the causes of behavior are as hidden to normal persons as they are to his nervous patients (The Basic Writings of Freud, ed. A. A. Brill (Random House, 1938), p. 954).

23 Freud, Five Lectures, p. 27.

24 Ibid., p. 50. Freud, so far as I know, does not explain how satisfying one desire — e.g., the desire to draw birds in flight — satisfies another, different desire — e.g., a sexual desire. He seems to think that the fact that one desire is a symbol for the other explains how such displacement is possible; but it is hard to see how one's hunger for food, e.g., can be satisfied by satisfying a desire to draw sheaves of wheat, even if the sheaves are a symbol for food.


26 Radcliffe-Brown, Structure, p. 144.

27 Ibid., p. 152.
The idea that the "natives" have beliefs about nature radically different from the anthropologist's own is itself a reason for the anthropologist to believe that she has misinterpreted the natives' beliefs. See Davidson's remarks on the role of charity in the interpretation of belief in Inquiries, essays 9–11.

Radcliffe-Brown, Structure, p. 142.


Ibid., p. 8.

Ibid., p. 206.

Ibid., p. 7.

Here Becker follows Milton Friedman, who argues in “Methodology,” that a theory is to be judged only by its predictive power for the class of phenomena which it is intended to explain and not on the reality of its assumptions. See chapter 5 for a discussion of the different interpretations of rational choice theory in economics.

For a fuller discussion of this distinction, see chapter 5.

It might be reasonable to believe that it is a law of nature that whenever A, B, in the absence of any understanding of how A causes B to happen — e.g., that ingesting aspirin reduces pain — but my point is that an understanding of the causes of B can be a reason to doubt that it is a law of nature if, according to that understanding, C is the cause of B, and there is no lawlike relation between A and C. For a more complete discussion of causes and laws in economics, see chapter 5.

Preference, according to some economists, is revealed preference, and revealed preference is nothing more than the internal consistency of observed choice. So understood, preference is not a state of mind or a possible cause of a person's choices at all, and if a participant's choices are consistent and she explains that she is acting contrary to her own preferences, she must be mistaken.


Insofar as the social sciences discount participant accounts; and, as I have tried to show, discounting is a widespread practice in the social sciences.

P. F. Strawson, Freedom and Resentment and Other Essays (Methuen, 1974), pp. 1–25.


My position is similar to that of the soft determinist who argues that it is not causal determinism that threatens moral responsibility but the denial that a subject's own reasons are the cause of her behavior. According to the soft determinist, subjects are responsible only if they are able to exercise causal control over their behavior, and such control is exercised when their behavior is caused by their intentions (rather than by external forces or internal compulsions). My position is not that a moral subject must have exercised causal control over a part of her behavior to be responsible for it, but that she must be able to exercise causal control over a large part of her behavior before she is responsible for any part of it.

An objection: my argument assumes that a person is rational only if her reasons are causes of her behavior; but some philosophers maintain that our concepts of rationality and causality belong to separate schemes. A person is rational, on their view, if she has beliefs and desires that are practical reasons for some large part of her behavior; but if her beliefs and desires rationalize her behavior, they do not cause it, for reasons cannot be causes. See, e.g., A. I. Melden, Free Action (Routledge and Kegan Paul, 1964), and R. S. Peters, The Concept of Motivation (Routledge and Kegan Paul, 1958). For a criticism of the view that reasons cannot be causes, see Davidson, Actions and Events, essay 1.

The objection, I believe, is without merit, because rationality and causality cannot be separated in the way that these philosophers suppose. Someone whose arm moves because it is pushed may have intended her arm to move to call attention to herself. Her intention is a reason for her arm moving; but because the push and not the intention is the cause of the movement, the movement is not intentional, and if not intentional, her intention does not explain by rationalizing her movement.

Freud, Five Lectures, p. 53.


The question has been much debated as to whether Freud's theory of double consciousness undermines moral agency or opposes otherwise appropriate moral evaluation. Those who maintain that it does argue that Freud's subjects, being the victims of the unconscious processes of repression or sublimation, are sufficiently ignorant of what they are doing and why they are doing it as to be blameless for doing it or are so coerced by these mechanisms to act contrary to their conscious intentions that their behavior is involuntary and so not subject to blame. For a useful survey of this debate, see M. W. Martin, Self-Deception and Morality (University of Kansas Press, 1986). My argument combines ignorance and coercion: Freud's subjects, on my view, suffer a forced ignorance of what they are doing and why they are doing it, and being ignorant and without the means to learn the truth, they are unable to exercise causal control over their behavior.

J. S. Mill, Utilitarianism, ed. Oskar Priest (Bobbs–Merrill Company, 1957), p. 10. Mill maintained that the quality, not merely the quantity, of satisfied desire is the measure of happiness. I ignore this here, for it does not affect the argument.


See Davidson, Inquiries, essays 9–11.

There are utilitarians who separate intentionality from moral worthiness. On their view, sentience, the ability to feel pleasure and pain, is all that is needed to be someone whose welfare matters. Utilitarians who believe that harm to animals is a moral concern on a par with harm to persons, and yet that the
behavior of animals is not intentional, do not assume, as I have suggested, that
only the bearers of intentions are members of the moral community. There is
some evidence that Mill himself thinks that the moral community includes “the
whole sentient creation.” See Utilitarianism, ch. 2, para. 10.
54 Ibid., pp. 57–8.
55 Ibid., p. 76.
57 Ibid., p. 9.
58 Ibid., p. 11.
59 Ibid., p. 18.
60 M. Frye, *The Politics of Reality: Essays in Feminist Theory* (Crossing Press,
1983), pp. 84–94.
61 Ibid., p. 84.
63 See chapter 1 for a discussion of the different reasons for the belief that our
basic institutions should be neutral on issues of the good; valuing autonomy is
only one.

The fact/value distinction

John Stuart Mill’s call for neutrality in the social sciences is based on
his belief that the languages of science and art are different. “Science
is a collection of truths; art, a body of rules, or directions for conduct.
The language of art is, Do this; Avoid that. Science takes cognisance of
a phenomenon, and endeavours to discover its law; art proposes to itself
an end, and looks out for means to effect it.” Like Hume before him,
Mill assumes that no body of rules (art) entails or is entailed by any
collection of truths (science). Most recent discussions of the role of values
in the social sciences rest on Mill’s assumption. Max Weber, in arguing
for value-freedom in the social sciences, for example, takes it for granted
that there can be a language of science – a collection of truths – that
excludes all value-judgments, rules, or directions for conduct. For Weber,
as I explain in chapter 2, a body of research or teaching in the social
sciences is value-free if and only if it does not contain any “practical
evaluations regarding the desirability or undesirability of social facts from
ethical, cultural or other points of view.”

In the preceding chapters, I argued that, given the way data are col-
lected and theorized in the social sciences, rules (values) cannot be excluded
from these sciences, but I did not question Mill’s assumption that there
is a logical gap between the languages of science and art. My point was
to show not that the social sciences cannot be value-free, Weber’s ideal,
but that they cannot be value-neutral – that is, neutral between competing
conceptions of the good – the broader liberal idea.

My aim in this chapter is to show that Mill’s assumption is mistaken
and that value-freedom in the social sciences is not possible. First, I argue
that proper attention to the open texture of meaning and the tie between
meaning and collateral information shows that words cannot be purged
of their evaluative content and that values cannot be eliminated from the
language of the social sciences. Second, I argue that proper attention to
the use of the language of the social sciences shows that whether speech