ROBERT NOZICK

ON AUSTRIAN METHODOLOGY

The major figures of the Austrian tradition in economic theory are Carl
Menger and Frederick von Weiser, originators of marginal utility theory,
Eugen von Böhm-Bawerk, and in this century Ludwig von Mises and the
co-winner of the 1974 Nobel Prize in Economics, Frederick Hayek.[1]

A framework of methodological principles and assumptions, which
economists in other traditions either do not accept or do not adhere to,
shapes and informs the substantive theory of Austrian economics. I shall
focus on the most fundamental features of this framework, the principle of
methodological individualism and the claim that economics is an *a priori*
science of human action, and upon two issues at the foundation of Austrian
theory within this framework: the nature of preference and its relationship to
action, and the basis of time-preference. I shall be forced to neglect the
farthest reaches of the theory, for example, the Austrian theory of the
business cycle, where still the fundamental methodological theses intertwine.
I also shall leave untouched other illuminating distinctive emphases and
approaches of Austrian theory, e.g. the constant awareness of and attention
to processes occurring in and through time, the study of the coordination of
actions and projects when information is decentralized, the realistic theory of
competitive processes. Nor shall I be able to detail the intricate inter-
connections of the different Austrian themes.

I. METHODOLOGICAL INDIVIDUALISM

The methodological individualist claims that all true theories of social science
are *reducible* to theories of individual human action, plus boundary
conditions specifying the conditions under which persons act.[2]

Methodological individualists are reductionists to the extent of their claim
that true theories of social science are reducible to theories of individual
human action, but typically Austrians *oppose* other reductionist claims, e.g.
that theories of human action are reducible to neurophysiology, chemistry,
and physics, or that social science is reducible to these in a way which

bypasses human action. This raises the question, which I shall not discuss here, of whether the *anti-reductionist* arguments which Austrian methodological individualists wish to use against the possibility of reduction from below (physics and neurophysiology) also can be wielded against methodological individualism itself, by nonindividualist social scientists who doubt the reducibility of social science to the level of individual human action.

We now must state the thesis of methodological individualism somewhat more precisely. Consider the question of whether we can reduce the theory of two-person interaction to the theory of individual human action. Economists who discuss individual human action often use the example of Robinson Crusoe, so we might call the theory of such individual action 'Robinson Crusoe theory'. Our question then is whether the theory of the interaction of Crusoe and Friday can be reduced to Crusoe theory.

Crusoe theory is the theory of Crusoe's interaction with the inanimate and nonhuman animate environment. Crusoe faces scarcity, allocates time and resources to some uses and foregoes others, does what he prefers, satisfies the principle of diminishing marginal utility, exhibits time-preference, saves from current consumption to increase future consumption, and so on. Crusoe theory will include the theory of individual decision under risk and uncertainty.[3] It will talk of various dimensions, such as actions, alternatives, expected consequences, estimates of likelihoods of expected consequences, and, I believe, expected utility, but nothing need be made of this now.

Does the theory of two-person interaction merely *specify* the previous Crusoe theory and apply it in particular circumstances, or does it introduce something new and irreducible? This question is interesting, and it is *not* clear what the answer is. Before discussing it further, I must contrast it with a trivial and uninteresting question and answer. Suppose it is asked whether the theory of two-person interaction is reducible to the theory comprised of all psychological truths about *individuals*, and that an affirmative answer is proposed, on the grounds that it *is* a psychological truth about Crusoe that when he interacts with another in a certain situation he behaves in a certain specified way.[4] This *is* a truth about Crusoe, but what kind of truth is it? In particular, is it a truth that *follows* from or is a specification of *the truths about Crusoe in noninteracting situations?*

Having distinguished the interesting from the trivial question, let us notice why the answer is not obvious. Crusoe can't just treat Friday's actions as

events of inanimate nature, form certain expectations about them, estimate the probabilities of the different things which Friday might do, and act accordingly and rationally. The problem is that what Friday will do in some situations depends upon *his* estimate of the probabilities of what Crusoe will do, where Friday does *not* merely treat Crusoe as an inanimate object, but as someone who also is estimating what he (Friday) will do. Each treats the other as a rational agent whose action depends upon (his estimate of) the rational course of action of the other, mutually realizing that the other's act depends upon his own in this way. Such situations are treated in the literature of game theory.[5]

Is the theory of action in game-theoretic situations reducible to, that is, derivable from the general theory of isolated action (Crusoe theory) *plus* a statement that the people are in, and (mutually) realize they're in a game-theoretic situation? Or, is the theory which is true of people in game-theoretic situations a new and irreducible theory? We cannot say merely that people will do what they think is best in this situation, taking this claim from Crusoe theory. The question is whether what they think best in *this* situation is determined by the Crusoe truths as applied to this situation. The answer to this question is not obvious, and we should note that theorists who work within game theory develop concepts (e.g. 'equilibrium strategy') which do not seem merely to be specifications of some notion(s) within the Robinson Crusoe theory. (That is, they do not seem to be merely notions of Crusoe theory restricted by parameters of two-person interaction.) To say that since all there is are individual people, the Robinson Crusoe theory *must* be sufficient, would be to answer the *trivial* question. ("It's true of him that in a game-theoretic situation he does such and such.") The interesting one would remain.

However, I do not wish to deny that the inanimate environment might mimic game-theoretic situations. For example, suppose Crusoe interacts with a mechanism whose states causally depend upon the actual reasoning Crusoe goes through. Crusoe's reasoning is mirrored in the states of his brain, and these states causally influence the mechanism whose motions then affect Crusoe. If Robinson Crusoe theory includes the theory of Crusoe's interactions with *this* bit of inanimate nature, it may well include game theory. Or perhaps without such a mechanism a person may (consider himself to) be in a choice situation wherein he must anticipate and perhaps thwart the choices of

his future self (who will not remember the earlier choice). So, again, the theory of one individual's behavior may come to include game-theoretic or similar principles.

It is plausible to try to demarcate such situations S as involving an agent's belief that outcomes are contingent (in a way which must be specified) upon how his environment reacts to own actions. Given this demarcation, Robinson Crusoe theory would be the theory of an isolated individual's behavior in non-S situations. On the other hand, if the demarcation cannot be drawn, so that by this route the theory of an individual's behavior includes game theory, Kantian moral considerations, etc., then the present epicycle about n-person interactions will not be needed in order to state the methodological individualist position.

Is the theory of 3-person interactions reducible to Crusoe theory *plus* the theory of 2-person interactions? Three-person situations involve the possibility of coalitions of two against one, of a member of the majority coalition being lured away into a new coalition with the outsider, and so forth. Suppose, for example, that $10 is to be divided up among three persons so that if (at least) two of the three stay agreed to a particular division, it gets instituted. Crusoe and Friday say, "let's each take $5, and freeze Defoe out". Defoe says to Crusoe, "I'll give you $6 and take only $4 for myself, so you're better off in a coalition with me than with Friday". Crusoe agrees. Friday then says to Defoe, "Let's divide the $10 by giving $5 to each of us, and freeze Crusoe out. That way we both benefit". And so on.[6] Once again, it is not obvious whether 3-person theory *is* reducible to 1- and 2-person theory, or whether instead new principles emerge which govern the new phenomena that are possible.

Our questions have been of the form: are the laws of $n + 1$-person interactions reducible to the laws of n-person (inter-)actions? The methodological individualist is committed, I believe, to saying

(a) that there is an n above which there is no change in the laws; all laws about interactions are reducible to the theory of n (or fewer)-person interactions; all social science is reducible to the laws ($\leqslant n$) of human interaction *plus* a specification of the situation (to which the laws are applied)[7]

and

(b) that n is *small*.

I admit that 'small' is an imprecise way of delimiting the claim. How small must n be, if methodological individualism is to hold true? Three? Anything less than ten? However, such imprecision does not undermine the contrast of methodological individualism with the view that, for example, a new theory (as opposed to a specification and application of the previous theory) is needed for the explanation of behavior of (and in) *crowds*.[8] The methodological individualist denies that there are specialized and irreducible psychological truths to the effect that a person in a *revolutionary* situation and conditions C does A, etc. It *is* logically possible that there be such emergent truths, irreducible to the *general* theory of n-person interaction.[9] (In this case, the completely general claim of the methodological individualist would be incorrect.)

How does the methodological individualist treat *institutions*? Let us take as an example (admittedly a favorable one) the institution of money. Menger offers us the following account.[10] To avoid the disadvantages which direct barter involves of not being able to trade for something you want from Crusoe who wants nothing you have, and therefore having to search for an intermediate Friday who has what Crusoe wants and wants what you have, people will tend to hold, trade for, and be willing to accept in exchange those goods they know others are (more) likely to accept in exchange; the more this is known, the more will such goods be traded for, and so there will be convergence upon a small number of goods which will, for obvious reasons, be initially valuable, portable, easily divisible in varying quantities, and homogeneous. Thus a medium of exchange precipitates out of the exchanges of individuals in a barter situation, each attempting to improve his own situation while taking account of the likely actions of others. Menger thus provides us with an explanation of the *creation* and of the *maintenance* via individual actions of a particular institution.[11]

Existing institutions also shape and affect the actions of individuals. They affect the opportunities available to people, and they shape people's utility functions. Furthermore, institutions are transformed and altered into different ones, and sometimes they are overthrown; this, too, is the result of individual actions. These statements are not very controversial. The methodological individualist adds that in explaining each such transition and affecting, it is a *general* theory (of $\leqslant n$-person interaction) which is specified and applied, and not some specialized theory which fits only that social situation. Our construal of the thesis of methodological individualism has

several virtues: under this construal, the thesis is an interesting one, it has real content, and could be false (emergence *could* be true), and at present its truth (I believe) is unknown. Furthermore, the explanations which methodological individualists offer, and view as satisfying their thesis, all fit the thesis so construed.[12]

Must methodological individualists speak of *institutions* at all? Rather, won't they view institutions merely as the *sum* of the actual acts done within them, so that any institution is nothing more than actual acts done by particular people (who occupy particular institutional roles but are otherwise identified)? However, institutions are not merely sums of particular act tokens, when if the person hadn't done those particular act tokens, he would have done other similar ones, or when if this person didn't occupy an office or role, another would have who would have acted similarly. In such a situation, the *subjunctives* must be explained also, not merely the indicative facts of which tokens were done when. It is the existence of such subjunctive facts which prevents the identification of an institution with the particular act-tokens done (as we say) under it. But though this identification is blocked, the methodological individualist can proceed to (try to) explain the subjunctives as well as the act tokens on the basis of previous act tokens, and so nothing about institutions need be left unexplained.[13]

Let us glance further at the patterns whereby actions create and maintain institutions which shape actions. An institution is self-sustaining if it shapes
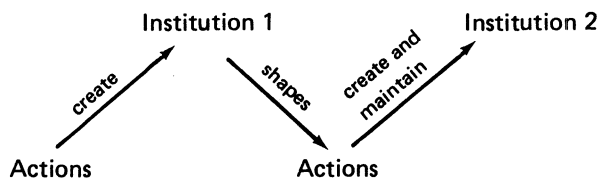
actions which maintain *it* (and are sufficient to overcome actions which tend to alter it). An institution is self-destructive if it shapes actions which alter and transform it.[14] There will be a regularity wherein Institution 1 is followed by Institution 2 if 1 shapes actions which create 2. Seeing this explanation will suggest what type of exceptions to the generalization there will be. The methodological individualist denies there is a law that Institution 2 must follow Institution 1, and that *because* of the law Institution 1 will

willy nilly shape some actions or other which lead to Institution 2. To explain the current situation of institutions and actions, we would show how current actions are shaped by yesterday's institutions and actions, and how today's institutions are yesterday's institutions altered by intervening actions (and maintained by some), with a cultural lag. Institutions don't disappear overnight, for they are embodied in modes of behavior which don't alter overnight.

If each time we explain the current situation as arising from previous actions *in* a certain institutional setting, then why are actions *prior*? In this apparent chicken and egg situation, why aren't we equally methodological institutionalists? Why think in terms of diagram 1 rather than diagram 2? Do we eventually get back to a starting point of actions and *no* institutions?

Institution   Institution

Actions   Actions

Diagram 1

Actions   Actions
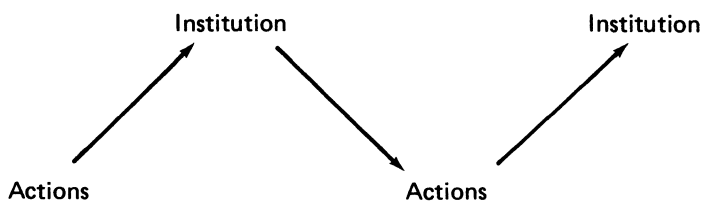
Institution   Institution
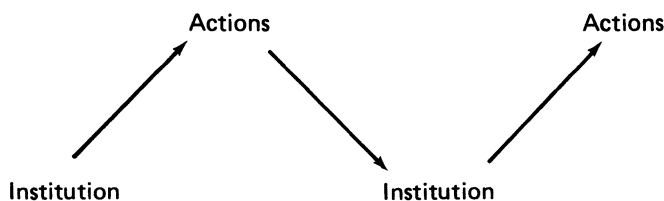
Diagram 2

The first people came from organisms with mating patterns, group relations, territorial patterns, and so on. We have no need to press the term 'institution.' Perhaps institutions are (by definition) transmitted and maintained *culturally*. Whereas these far past patterns were (initially) passed on *biologically*, as a result of evolutionary selection.[15] Given these facts, does the methodological individualist win or lose? He wins, I think, if he can explain all with the theory of human action plus evolutionary biology. *Institutions* are then, initially, a dependent variable.

However, the process of evolution may have instilled desires which themselves refer to institutional or particular social situations. In this case, even though non-macro-social explanations can be offered *from the beginning*, social scientific explanations of current behavior would have to admit (innate) desires or reinforcers for which macro-social reference would be needed to specify either what is desired or the conditions under which the desire is operative. The strongest and most general thesis of methodological individualism would then be false, and though the facts mentioned would surely not greatly perturb the methodological individualist, it is difficult to see how to modify his thesis so as to make it compatible with such biological facts and similar ones as yet unknown, while still avoiding the *trivial* thesis discussed earlier.[16]

We haven't yet focussed upon the part of the scheme where institutions shape actions. The methodological individualist, we have seen, is distinguished by his claim that it is a *general* theory of action which gets applied, not particular theories which are irreducible to general laws of action. But what are the general laws whereby institutions shape actions, what are the general laws of the shaping of utility functions? The Austrian tradition has devoted little attention to this question, perhaps because it was thought that all substantive features of utility functions (or preferences) were matters either of biology or of free choice. If methodological individualism is true, there will not be irreducible laws of the form: people brought up under institution I (tend to) have such and such a type of utility function. The fundamental laws will not mention a *particular* kind of institution; rather there will be general laws of utility shaping, which are then applied to the situation of a particular institution, in order to show how *it* shapes. Institutions (the stable patterns of others' actions) provide opportunities for doing various things, and a certain patterning of rewards and punishments for actions. One general theory of the shaping of utility functions might state how opportunities, rewards and punishments, and contingencies of reinforcement, combine to shape utility functions and preferences. The framework of *this* sort of theory is provided by the theory of operant conditioning.[17] This theory has been much attacked by libertarian writers, but it is important to see that methodological individualism requires *some general* theory (it need not be this one) of how utility functions are shaped in institutional environments.

Supposing the thesis of methodological individualism is true, what are its

consequences for the practice of social science? The fact that all true theories of social science are reducible to general theories of human action and interaction ($\leq n$) does *not* mean that all true theories already will have been reduced, or that it will be obvious how the reduction is to be carried through. Therefore, a social scientific theory formulated at the macrolevel *cannot* be condemned merely because it is (as yet) unreduced. It could be condemned (supposing the thesis of methodological individualism were known to be true) if it were known that it was *impossible* to reduce it to theories of human action. But it is doubtful in the area of social science that a proof of the irreducibility of some theory will be forthcoming. (Though perhaps one could show that a *contrary* theory at the macrolevel follows from a specific consistent theory of human action.) Does it *follow* from the thesis of methodological individualism that the *proper* way to construct a social theory is to *start* with the theory of human action, and to work one's way up *from it*? (This is the procedure of the Austrians.) Compare the situation of someone who believes that the true theories of biology are (ultimately) reducible to physics and chemistry. He will not condemn the biologist for failing to do physics and chemistry, and he will not condemn biological theories which have not yet been reduced (even though he does believe that whatever true biological theories are discovered *will* be reduced to physics and chemistry *eventually*.) Social scientists should be the last to claim that there cannot be a similar division of labor in *their* intellectual world. Thus, it appears that no consequences *need* follow from the thesis of methodological individualism about what our attitude should be to any given as yet unreduced macrotheory of social science. Methodological individualists, like everyone else, will have to assess the truth of such theories by examining the evidence and arguments for and against them. This conclusion is so much at variance with the usual Austrian methodological individualist view that it suggests an attempt be made to formulate the thesis of methodological individualism so that it is *not* merely about the reducibility of theories. If I knew how else to formulate it, I would.

II. THE *A PRIORI* SCIENCE OF HUMAN ACTION AND ITS APPLICATIONS

One branch of Austrian theorists (Mises, Rothbard) holds that (many of) the

most important general laws of human-*n*-person interaction are derived within (as they term it) an *a priori* science of human action. This science, it is held, begins with an examination of the essence of action; that is, purposeful behavior and, in elaborating what is contained in this essence, is led to a body of necessary truths holding of human action as such.

Note that holding there is this *a priori* science of human action is logically independent of accepting the thesis of methodological individualism. For a methodological individualist might hold that all true social scientific theories are reducible to laws of human action, but that *none* of these latter laws are necessary truths. And someone might hold that there is a body of necessary truths about human action, but that these do *not* suffice to reduce and explain all true theories of social science[18]

The idea of elaborating what is contained in the essence of human action is certainly an interesting and challenging one. Such a project, involving as it does synthetic necessary truths, would have been ruled out by the logical positivists as impossible and empty. But the positivist position and arguments on synthetic necessary statements, as much else within the positivist position (for example, the verifiability criterion of meaning), has fallen upon hard times. Recently, there has been a resurgence of interest in such statements, and it is fair to say, I think, that there are no arguments generally acknowledged to be compelling against the possibility of such synthetic necessary truths.[19]

This is not the appropriate occasion to investigate the general possibility of such truths. Instead, we shall consider the implications for the practice of economics if there were such a body of truths. Later sections of this essay will consider particular such (purported) truths. I shall consider only some of the (purported) derivations from the essence of action, but I should mention that there is a need for a clear, precise, and consistent statement of the content of the (*a priori*) theory within a specified vocabulary of primitive terms and with explicit definitions and axioms. Such a statement should make clear whether preference is initially over actions or outcomes; it would fix one notion (preference, satisfaction, desire) as primitive and define the others in terms of it; it would be sophisticated enough to take account of the considerations of the theory of choice under uncertainty and consistently distinguish in the presentation of the theory what it is expected *will* happen from what it is thought *might* happen; and it would avoid Mises' unfortunate

tendency to speak as if the outcome of an action is preferred to the current situation (it need not be) rather than to what *would* obtain if the action weren't done. More care also is needed in stating the future orientation of action, for the point of an act may merely be to do it, or to be continuing a previously started plan, or to be following a previous commitment. Thus, it is a mistake, I think, to speak as Mises does of acting man necessarily ignoring sunk costs. It may be *irrational* to consider them, but people in restaurants, for example, whose food has come and turned out to be poor tasting certainly often *speak* as though the reason they're going ahead to eat it is that money already has been committed to it. True, one might explain their sticking with this food rather than ordering another (preprepared) dish in the same way we explain why the newest, most modern equipment is not immediately installed in every factory. But, on the other hand, it is not *impossible* that letting something he's paid for go uneaten has disutility for that person. Economic theory *can* assume entrepreneurs ignore sunk costs, not because doing so is part of the essence of human action, but because the market tends to filter out entrepreneurs who take account of sunk costs. Finally, I would want the theory to be formulated so that even though preferring is a subjective psychological state, the ultimate things which are preferred one to another *need not* themselves be subjective psychological states (such as felt satisfactions or dissatisfactions, or removals of such things). This, I realize, is a more complicated question, but I need not linger over it here, since I have stated relevant considerations elsewhere.[20]

I must make only these brief remarks about formulating the theory, since I wish to examine in somewhat more detail the connection of preference, choice, and action, and the theory of time-preference. But first, in this section, consideration must be given to the *application* of a body of necessary truths about action. Actions are a certain type of behavior, namely, purposive behavior. Not all behavior is purposive. There is caused bodily movement which is not even behavior, such as the movement of a body when dropped unconscious from a height. There is behavior which is not action; for example, simple unconditioned reflexes. By training, such behavior might come under the control of choice, but if such training didn't take place, the reflex behavior wasn't action (though the earlier 'nontraining' may have been). What about the conditional reflexes of classical Pavlovian conditioning? The ringing of the bell has been paired with food, and now (with no food present)

the bell rings, and the dog, or the human being, salivates. Is this salivating behavior an action? I do not recall a discussion specifically on this point in the writings of the Austrians, but their answer, I believe, would be that such Pavlovian reflexes are not actions.

More interesting and difficult is *operant* behavior. Is behavior which falls under the principles of operant conditioning, action? Processes of operant conditioning by use of rewards and punishments, positive and negative reinforcements, on various contingencies of reinforcement, raise and lower the probability of various behaviors, and introduce various patterns of extinction[21] Is such behavior which is subject to and shaped by reinforcement, action? Austrians might quarrel with applying principles of operant conditioning to people and to theorizing about people in this way, for the behavior the Skinnerian psychologist applies his principles to is the very behavior the Austrian theorist would call 'purposeful,' that is, 'action.' The Skinnerian psychologist applies his principles to animals as well, which raises the question of whether the Austrian theorist counts operant behavior done by, e.g., rats and pigeons, as action. Again, I do not recall a specific discussion of this question by an Austrian theorist, but I would think they would be reluctant to apply to animals many of the categories that go with actions, e.g. expectations, plans, image of a desired end. Recall that principles of operant conditioning have been observed to operate quite low on the evolutionary scale.

If such operant behavior of these other organisms is not action, then we have an interesting example of a type of behavior to which the categories of human action do not apply. The principles of operant conditioning, let us assume, apply to these other organisms. (Note that there is an evolutionary account of the adaptive advantage of operant conditioning; being operantly conditionable would be selected for in the process of evolution among organisms that were merely subject to classical conditioning.) Might those principles also be true of us, of large segments of the behavior of human beings, as an *a posteriori* theory? Do we know *a priori* that its categories and statements don't apply to much of human behavior?

There *is* this alternative, relatively elaborate conceptual scheme which does not appear to talk of or within the Misesian categories, yet whether it is true of people or not appears to be an empirical question, to be decided *a posteriori*. (Did these principles just stop being true at a certain point on the

evolutionary scale? Does human consciousness and intelligence provide new means through which these principles work, or instead an area in which they don't work at all?) It may be that Skinnerian theory and human action theory are compatible, and can both be true of the same behavior. If so, even if human action theory *is* known *a priori*, it may be that it is not the best theory, for the Skinnerian theory actually may explain more, enable us to predict more, and so forth. In this case, the possession of an *a priori* theory would *not* stop us from pursuing and developing, formulating and testing, a body of contingent truths about behavior. Possession of an *a priori* theory cannot tell us that there won't also be a better, more predictive, wider ranging *a posteriori* theory, and so *cannot* bar the pursuit of empirical science. (This point, obviously, does not depend on anything about Skinnerian theory, but it seemed useful to make it by considering an actual alternative framework, very different from Mises', which functions within an *a posteriori* theory currently under active investigation.)

On the other hand, it may be that the two theories cannot both be true of the same behavior, that the truth of one excludes the truth of the other. Many Austrians write as though there is *no* set of concepts, other than theirs, via which one conceivably might understand much of human behavior. The existence of Skinnerian theory casts this in doubt, and supposing this theory to be incompatible with human action theory, it raises the delicate question of which of the two bodies of theory is true of most of our behavior in the world. (In Austrian terms, it raises the question of whether that behavior *is* action.)

Note the importance for the Austrian view that there be nothing *else* close to action, nothing very similar to it. On that assumption, human behavior which isn't action can easily be distinguished. It would not be a delicate matter involving much observation and testing to decide *which* theory holds true of some behavior which appears to be action. On the other hand, suppose there were behaviors similar enough to human actions so that they could not be distinguished at one or even two glances. Suppose further that such nonaction behavior occurred frequently, often intermixed with the actions of persons. The nonaction behavior *does* obey different laws, though, e.g., the laws of operant conditioning. Even if there were an *a priori* theory of human action, the question of whether this theory fit a particular situation would have to be determined by detailed empirical investigation, complicated

statistical procedures, etc. That is, the claim that the theory of human action fits a situation would require as much evidence as the claim that some other *a posteriori* theory fits a situation. The *a priori* character of the theory of human action would be of *no help* in deciding whether it was true of a behavioral situation, and economics would be carried on in as empirical a manner as any other science. (Even if somehow in this situation someone could know *a priori* that it was an action he was performing, others couldn't know this *a priori*.)

Thus, it is important for the Austrian view of how economics is to be carried on to add to their claim that there is an *a priori* science of human action, the claim that there is a great enough gulf between human action and other human behaviors to make it easy to identify which behavior is action. Otherwise, the only way to determine whether or not behavior is action might be to determine which laws it satisfies. And this might require as much empirical investigation as goes on in an *a posteriori* science of human behavior. Can the existence of such a gulf be known *a priori*; is it a necessary truth that no behavior mimics action closely enough to require detailed investigation to separate them? I do not see that it is, and the existence of an *a posteriori* theory of this sort (e.g., Skinnerian theory) has the consequence that the science of human behavior must, to a great extent, be carried on as an empirical science. Even if it were known *a priori* that actions satisfied certain laws, it would be an empirical question whether particular bits of behavior are actions and, hence, whether they satisfied those laws[22]

Does the method of *verstehen* give us knowledge of which conceptual structure applies to people's behavior?[23] Critics of *verstehen* have claimed that it is at best a route to thinking up hypotheses, and not a way of coming to know which hypotheses are true.[24] Current views on the testing of hypotheses describe a role other than the suggesting of hypotheses or the showing of which one is true, namely, the assignment of plausibilities to hypotheses or (on the Bayesian view of statistics) of prior probabilities to these hypotheses. Such plausibilities or prior probabilities will affect what it takes to show a hypothesis is true or false. In an unpublished essay, Hilary Putnam makes the interesting suggestion that the process of *verstehen* comes into this prior assignment, and so legitimately plays a role (though not a *conclusive* one) in the process of accepting and rejecting hypotheses.

Even if we grant Putnam's point, we should step back to examine what

might underlie the use of *verstehen*. Consider the enterprise of what, following Quine's use of the term 'radical translation'[25] we might call *radical verstehen*, the understanding of the behavior of organisms of another species or with whom we have *no* biological links since they evolved on another planet. Can we *verstehen* such organisms? Empathetic understanding might suggest hypotheses about their behavior, but these would have to be tested further. The *general hypothesis* that we are *able* to *verstehen* their behavior would be tested in the process of testing the specific hypotheses about them which empathetic understanding suggests.[26] What is clear is that we could not know *a priori* that we have the ability to empathetically understand the behavior of these organisms. If we do have this ability for those organisms, we will have to come to know this *a posteriori*, presumably by appropriate procedures of psychological science. It might turn out that we have this ability with only *some* types of their behavior; though empathetic understanding suggests with equal plausibility hypotheses about other types of their behavior, further testing might show our success rate with these latter hypotheses to be no better than random. Our reliability with different types of their behavior will be determined *a posteriori*.

Closer to home, consider the claim that *you* can *verstehen* the behavior of other persons. That, too, it seems, is not known *a priori*. Presumably, you will have confidence (if you do) in your ability to so understand the behavior of others because those of your everyday attempts to do so which had further predictive consequences have borne up pretty well. But such a nonrigorous test of the hypothesis cannot show that an alternative conceptual scheme is inapplicable when this alternative scheme often would have roughly the same further predictive consequences about the behavior of others in the situations of everyday life. In that case, the apparent success of our episodes of empathetic understanding might be only apparent.

But can't we know, without intricate testing, which concepts apply to our *own behavior*, that is, can't I know which apply to mine, and you know which apply to yours? We each certainly do apply a particular conceptual scheme to our own behavior with great confidence.[27] But on what basis do you do so and, in particular, do you know *a priori* that this scheme applies to you? Perhaps one wants to say that you know which concepts apply to your behavior since you're *doing* it, after all, and doing it in a way which *infuses* those concepts into the behavior which you produce, and which also leads

you to know that those concepts are so infused. You know you're making it happen *as an action*. As it stands, this is too murky to be helpful. Clarifying it is unlikely to show that the way the concepts get infused into behavior you produce while under the sway of the concepts, makes the theory utilizing *those* concepts the best explanatory theory of your behavior. (And there still would remain the problem of knowing what concepts to apply to the behavior of *other* persons.)

Still, there is the fact that we *do* apply, with confidence, the categories of human action. Another theory, e.g., the Skinnerian theory, will fail to generally undermine this confidence until it is combined with a theory of concept formation so as, from its own perspective, to explain the (cross-cultural) fact that people confidentially hold to the usual (non-Skinnerian) categories. If the Skinnerian theory could explain why, though the usual theory is (as it claims) false, almost all people adhere to it, this would make it more difficult to retain undiminished confidence in either the usual theory or in the process of *verstehen*, which uses its categories. However, even in this case it would not be impossible to continue to believe the usual theory, and the Skinnerian theory itself may predict this will happen, since belief in the usual theory was reinforced on a schedule which makes it highly resistant to extinction. Indeed, we may question how faithfully operant conditioning theorists adhere to their own theory when they try to argue the public away from the product of long and varied processes of reinforcement!

Verstehen cannot, we have seen, help to eliminate all empirical questions about the application of Mises' or any other theory of human action. However, we should not leave the subject of *verstehen* without remarking that it *is* possible that people (or some people) do possess the capacity of empathetic understanding, for at least some types of behavior, and that this capacity might be trained so as to become even more reliable. Whether this is so or not is an empirical question. Suppose that investigation by the procedures of empirical science identifies some persons as especially reliable *verstehers* of at least some types of behavior; as it turns out, they rarely or never are inaccurate. Even if we did not possess a theory of *how* this capacity operated, we might use such reliable *verstehers* as *instruments* by which to learn various things, even as we use nonhuman detectors of facts. Tests might yield the result that particular *verstehers* had a reliability as high as that we attribute to our usual emprical procedures of theory construction and testing, and in that case we would no more hesitate to say that *verstehen* has given us

knowledge than to say that the rest of science has. The standard procedures of the rest of science would be what established the reliability of *verstehen*, but once this *were* shown, there is the possibility that social science might then be carried on very differently.

One final remark about the applicability of *a priori* laws. Questions about this arise due to considerations *within* the theory as well, for example, for the principle of diminishing marginal utility. It is often objected to this principle that an additional unit of X (say, the sixth) may have *increased* utility since, now, along with the previous five units, the person is able to acquire something previously unavailable. The Austrian reply points out that in this case the relevant size of the unit is $6X$; the principle is meant to apply to situations where each additional unit of X can acquire only those things which previous units also were able to bring (and together with the others can bring only a larger conjunction of previously available things.)[28] The *size* of the units must be chosen so this condition is satisfied. Now, it is logically possible, of course, that for *any* choice of a finite size of the unit X, *whenever* a new nth unit of X becomes available, there is also a new *total* use of the n X's (that is, of the new X and the previous ones together) which ranks higher in the person's preferences than the sum of any n of the previously available things (and than the sum of the conjunction of something obtainable with one X along with anything obtainable with $n-1$ X's together, etc.) Let us call such a situation an *expanding universe*. A person's universe is *locally expanding* if it is an expanding universe over as many units as he has; that is, for any choice of the size of unit, so that he has at least one. The principle of diminishing marginal utility is not *false* of a person in a locally expanding universe; it just doesn't *apply* to him. The combination of his preferences and the opportunities available to him does not satisfy the condition of application of the principle. Since whether or not a person is in a locally expanding universe is an empirical question, it is an empirical question whether or not the principle of diminishing marginal utility *applies* to a particular person in a particular situation.

III. PREFERENCE, CHOICE AND ACTION

Austrian writers put forward a number of interconnected theses about the relations of preference, choice, and action:

(1) If a person does an action A, then that person preferred doing A to doing any other act which (he believed) was available to him at the time.

(2) No evidence can establish that a person prefers A to B in the face of a choice of B when A was available to the person.

(3) The notion of preference makes no sense apart from an actual choice made.[29]

Does all action show preference? (That is, is 1 above true?) Mightn't the person be *indifferent* between what he did and some other alternative available to him? We might have the following picture: of the acts available to a person, some he is willing to do, and others he is unwilling to do. He prefers A to B if and only if in a choice between A and B he is willing to do A and unwilling to do B. Let us grant that if he does A, this shows he was willing to do A. But doing A does *not* show he was unwilling to do B. He might have been willing to do B also. Thus, doing A does not show the person preferred doing it to doing B.

Writers sometimes speak of *weak preference*, of the person weakly preferring doing one act to doing another. We can understand 'the person weakly prefers doing A to doing B' as: the person prefers doing A to doing B, or the person is indifferent between doing A and doing B. In terms of this relation of weak preference, a person is *indifferent* between doing A and doing B if and only if he weakly prefers A to B, *and* he weakly prefers B to A. A person *strongly* prefers A to B if and only if he weakly prefers A to B and he does not weakly prefer B to A. Choosing an act may well be a sufficient condition for weakly preferring doing it to doing one of its alternatives. But it is not a necessary condition for that and, hence, it is not a sufficient condition for *strong* preference. There are other indications of indifference; for example, the person flips a coin between doing A and doing B and acts on the outcome of the flip; or, the person uses a random device to choose between doing A, doing B, and flipping a coin to decide between doing A and doing B, and so forth.

Indeed, the Austrian theorists *need* the notion of indifference to explain and mark off the notion of a commodity, and of a *unit* of a commodity. If everyone or one person prefers one homogeneous batch of a stuff to another homogeneous batch of the same shape of the same stuff (perhaps they like to choose the left-hand one, or the one mined first), these are *not* the *same* commodity. They will have different prices. Particular things x and y will be

the *same* commodity (belong to the same commodity class) only if all persons are indifferent between x and y. Without the notion of indifference, and, hence, of an equivalence class of things, we cannot have the notion of a commodity, or of a unit of a commodity; without the notion of a unit ("an interchangeable unit") of a commodity, we have no way to state the law of (diminishing) marginal utility.[30]

Does the existence of behavior, when a choice was made, show *what* was chosen and what was preferred? The behavior done does not show *to what* something was preferred, for it does not show what behavior was *rejected*. That depends upon the person's *beliefs* about what alternatives were available to him, and these are not shown by his actual behavior. We might think it is shown by his talk, but why can talk show beliefs about the alternatives available yet not show preferences? We should not answer, "because beliefs issue in talk and preferences issue in action." For, don't beliefs sometimes issue in action, and preferences in talk? Rather, aren't talk and action each the product of beliefs *and* preferences *both*?

Do we at least know *what* was the preferred alternative? Note that the behavior performed can be described differently, e.g., 'travelling to the first floor,' 'stepping in this particular place,' 'exerting a certain pressure.' *What* did the person prefer doing to something else? Preference, it seems, does not apply to things directly, but applies only via *descriptions*, via something linguistic-like. We cannot just say "*that* behavior," and point. For *which* is that? We might try to narrow things down by noting that what he preferred was something he *knew* about; hence, the description has to be one he knew of at the time he was acting or choosing. But which of these descriptions he knows of is connected with his preference? Perhaps the answer is, *all* that he knows of: he prefers the combination of all the descriptions he knows of for the behavior he does, to the combination of all the descriptions he knows of for each of the alternative behaviors he doesn't do. But he may have *neglected* and even repressed many descriptions he knows of. He knows of many he didn't consciously think of at any point in the process of deciding; he thought of some which he forgot, but could recall with effort, and so forth.

It might be suggested, instead, that he prefers the combination of descriptions that (he thought) fit the behavior done which he was actually thinking of, as he decided, to the combination he was then actually thinking

of which (he thought) fit the behavior rejected. But mightn't he then have been thinking only of the *worst* description, and then decided, 'I'll do it nonetheless"? Need he have been thinking then of a description at all? Couldn't he have gone through deliberations, and then when the time came, just acted without consciously thinking about it?

What is preferred, it seems, is hidden from an observer. The most an observer can say, for sure, is that when action takes place something or other is (weakly) preferred to something or other else; more fully, something or other connected with the behavior done was (weakly) preferred to something or other else not connected (as closely) with the behavior done. It is often said that an external observer cannot *predict* behavior on the basis of preference. Here we see that he cannot retrospectively *describe* preference, even on the basis of the *behavior* done!

Must it be, at least, that the person who acts knows what he himself preferred? Can't he say, "I thought about many things about what to do, and then just acted later without thinking about it then. I don't know what about it I preferred"? Is this impossible? or just rare? "But he knows he preferred doing it!" Need he *know* what the *it* is which he preferred? "Well, he *did it*!" True, but who is denying that?

Let us now turn to the strongest of the above statements about the connection of preference and action, viz. (3): the notion of preference does not make sense apart from an *actual action* which exhibits it. Notice that this is stronger than saying that preference is *somehow* connected to choices, to what a person *would* choose under certain circumstances. It says the notion is inextricably tied to *actual* choices.

It is not clear that the Austrians can consistently maintain this strong thesis. Consider, for example, Mises' view of action as exchange.[31] An action *A* is done instead of another act *B*. What is not done, *B*, is the *price* of doing *A*. The value of *B* is the *cost* of doing *A*. Mises then goes on to talk of the profit of an act, and also of the (I assume, *ex post*) loss. Now the cost of doing act *A* cannot be the value of *everything* foregone, that is, it cannot be the *sum* of everything foregone. If it were, we almost never would have profit. Suppose there are five alternative acts, closely bunched, and you do the one act, *A*, which is the best. The cost of *A* is not the *sum* of the values of the other four. (This would greatly outweigh the value of *A*.) The cost of *A* is the value of the *best one* of the rejected alternatives. The best one of the

rejected alternatives is the one you strongly prefer to all the others, or (if there's a tie at the top of the rejected alternatives) one you weakly prefer to each other rejected alternative. Notice that we are speaking of a person's preference apart from any *actual choice* of the preferred alternative. The next best alternative to A, call it B, is preferred to any of the others C, D, E, *though B itself is not chosen.* If we are to speak of the cost of A, when there is more than one other alternative rejected, it must *make sense* to speak of preference apart from an actual choice or doing of the preferred alternative. If *that* doesn't make sense, then neither does the notion of the *cost* of the action which was actually chosen.[32]

How *can* it make sense to speak of preference without an actual choice? Granting that the Austrians are wrong in denying that it makes sense, how is it possible? Let me present a sketch of a plausible account. Preference is connected with a *subjunctive*: to say a person prefers A to B at a time is to say he *would choose A* over B if he *were* given a choice between (only) A and B at that time. There's something about him in virtue of which he would choose A over B. It is not, for example, a random event.

That an event actually happens does not show that a subjunctive is true. Consider photons emitted toward a screen with two slits, and suppose that it is truly a random event which slit any photon goes through. If two people bet on which the next photon *will* go through, one saying it's one slit and the other saying the other, the winner is the one who picked the slit which turned out to be correct. The indicative statement is shown true by the event. In contrast, if someone says, "if a photon *were* emitted in the next second, it *would* go through the right-hand slit," and one is emitted, and does go through the right-hand slit, this does not show he was right. The *subjunctive* statement is *not* shown true by the event.

Perhaps this can be made intuitively more obvious via a recent account of subjunctives. Under this account, a subjunctive of the form 'if p were true, q would be true' is itself true if and only if in all those possible worlds where 'p' holds true which are *closest* to the actual world, 'q' also holds true. The minimal change from the actual world necessary to make p true, brings the truth of q along with it (in those closest worlds).[33]

Now it might turn out that those closest worlds in which p is true are neither uniformly q nor uniformly not-q. In that case, in the actual world neither of the subjunctives 'if p were true q would be true.' and 'if p were

true, not-q would be true' will hold true. If it is truly a random matter which slit the photon goes through, then before the photon is emitted, for each possible world in which a photon is emitted and goes through the right (left) slit there is another possible world, equally close to the actual one, in which a photon is emitted and goes through the left (right) slit. More precisely, the "photon is emitted" — neighborhood of the actual world contains worlds where the photon goes through the right slit, and it contains worlds where the photon goes through the left slit. The subjunctive 'if a photon were emitted it would go through that slit' is not true.

To say a person did A rather than B because he preferred A to B, is to say he did A rather than B because it's true of him that he *would* choose A over B. This is not a very potent explanation, but it does put the action in a pattern; it says the action flows from some underlying *disposition* (which might be very fleeting) to choose A over B. The occurrence of an action does not demonstrate that the subjunctive is true; also the person might have been indifferent. But often an action will stem from an underlying preference and will indicate that there is such a preference.

On this view of preference as constituted by the subjunctive to the effect that a person *would* choose A over B if he *were* given a choice between the two, we can see how to make sense of preference without an actual choice of the preferred act. It might be that a person would choose A over B in a choice between the two, but since he didn't have to make a choice between the two, that preference was not exercised; he did not actually choose A over B, and he didn't do A.

A substance is *soluble* in water if it *would dissolve* if placed in water. Substances which are never placed in water may yet be water-soluble, even though they never actually dissolve. 'Prefers A to B' is like 'soluble'; 'chooses A over B' is like 'dissolves.' The claim that it makes no sense to say a person prefers A to B unless he's actually chosen A over B, is like the claim that it makes no sense to say something is soluble unless it already has actually dissolved. Both claims are mistaken.[34]

The cost of an action, we saw, is the value of the most preferred alternative not chosen. To say that some unchosen alternative is preferred to every other one is to say it *would* be chosen over each other one in pairwise choices. And this might well be true. If the Austrians were correct in talking of scales of values as existing only in actual choices, there couldn't be a particular cost of a choice.

Subjunctives give us the possibility of making sense of *rationality* conditions on preference, e.g., the condition that preference be transitive. To say that preference is transitive at a time t is to say that for any three acts X, Y, and Z if

(1) it's true that the person would choose X over Y if given a choice at time t between (only) X and Y

and

(2) it's true that the person would choose Y over Z if given a choice at time t between (only) Y and Z

then

(3) it's true that the person would choose X over Z if given a choice at time t between (only) X and Z.

It is possible that (1) and (2) are both true for some X, Y, and Z, yet (3) is not; rather, (4) might hold:

(4) it's true that the person would choose Z over X if given a choice at time t between (only) X and Z.

Also, it is possible that *no* subjunctive, neither (3) nor (4), is true of the choice between X and Z. In this case, also, preference fails to be transitive. It is possible that (4) is true along with (1) and (2), because each of these subjunctives talks about different hypothetical circumstances. Since (1) and (2) do not talk about a choice between (only) X and Z, they don't, even together, settle what *would* happen under that condition. However, they arguably do settle what should happen, and so we have room for a rationality condition.[35]

Mises considers and objects to the requirement that preference be transitive.

The attempt has been made to attain the notion of a nonrational action by this reasoning: If a is preferred to b and b to c, logically a should be preferred to c. But if actually c is preferred to a, we are faced with a mode of acting to which we cannot ascribe consistency and rationality. This reasoning disregards the fact that two acts of an individual can never be synchronous. If in one action a is preferred to b and in another action b to c, it is, however short the interval between the two actions may be, not permissible to construct a uniform scale of value in which a precedes b and b precedes c. Nor is it permissible to consider a later third action as coincident with the two previous actions. All that the example proves is that value judgments are not immutable and that

therefore a scale of value, which is abstracted from various, necessarily nonsynchronous actions of an individual, may be self-contradictory. (*Human Action*, p. 103.)

The acts cannot be synchronous, but the subjunctives *can* hold true at the same time. So we can make sense of nontransitive preferences at a time. (Still, it does not follow that the *action* is irrational, only that the structure consisting of that action *plus* those preferences violates a normative requirement. *Which* should be changed is left open.) Leaving preferences aside, what are we to say about the corresponding nontransitive subjunctives? Isn't someone irrational of whom (1), (3) and (4) hold true?[36]

If preference is specified by choices that would be made in pairwise choice situations, there is room for normative conditions in addition to the transitivity one. For there will be conditions connecting pairwise choices with nonpairwise choices. For example,

> If, if there were a choice only between X and Y, then X would be chosen, then if there were a choice between X, Y and some other alternatives, then Y would not be chosen.

and

> If in a choice among the members of set S, X would be chosen then in a choice among the members of a subset T of S of which X is a member, X would also be chosen.[37]

Why does Mises think it so important to argue that the structure of preferences cannot be irrational? Perhaps because he doesn't want anyone interfering with choices on the grounds that they arise from irrationally structured preferences. That, however, is another issue, and one might think it possible for people to have irrationally structured preferences and also hold that this is within their rights, so that they may not be interfered with on the grounds that their preferences are irrationally structured (unless they have explicitly contracted into such a scheme of interference). I should not neglect to mention that the view of preferences as embodying (and embodied in) subjunctives about choice opens up the possibility of formulating further conditions on preference among probability mixtures of outcomes (in the manner of Von Neumann and Morgenstern) or among uncertain actions (mappings of states of the world onto consequences, in the manner of L. J. Savage) so as to yield measurement of utility on an interval scale.[38]

Of the three statements connecting preference, choice, and action which we listed above, we have argued that the first and third are incorrect. What about the second? Can evidence establish a preference of A over B in the face of an opposite choice of B, when A was available to the person? Can there be indirect evidence that a person would choose A over B if given a choice (only) between A and B, so that we conclude he *does* prefer A to B, despite the fact that he has chosen B when A was available? The person might have chosen B over A and over lots of other things, over so many things, in fact, that he lost sight of A. Here would be a case where he would choose A in a choice (only) among A and B, yet in a choice among A, B, and many other things, he would choose B, because he would lose track of A. Or perhaps there's no "would" about the matter; he just gets confused and acts; no *subjunctive* is true of the wide choice. The person actually chose and did B (violating one of the rationality conditions listed earlier), yet this doesn't show he didn't prefer A to B. We might stand by the claim that he does prefer A to B, offering a special explanation of why he chose B that time.

These examples, which are designed to cast doubt on (2) above, involve choice among a large number of alternatives. There cannot be a similar counterexample which involves only a pairwise choice, if preference goes with a subjunctive involving "would *always* do A rather than B." If, however, preference goes with a subjunctive involving "would . . . unless . . ." then there is the possibility of a pairwise counterexample to (2). There remains, however, another difficulty with (2)'s claim that no evidence can establish that a person prefers A to B in the face of a choice of B when A was available to him. From a person's doing B, we cannot know he believed A was available to him, or how he conceived the rejected alternatives. If our earlier discussion of whether we can know *to what* something was preferred was correct, we cannot possibly, no matter what we conclude, fly in the fact of a choice of B over A. For that is never a fact which *faces* us.

Rationality conditions are conditions which it is possible to violate. In contrast, it is difficult to see how preference *could* fail to be irreflexive and asymmetric. These seem to be part of the notion of preference, and it is a virtue of the subjunctive account that it makes no sense of instances where preference is reflexive or symmetric. However, though (in contrast to constitutive conditions) rationality conditions *are* conditions it is possible to violate, they had better not be violated *too much*, without special

explanation. For if some organism violated them just about always, it would be unclear that it was *preferences* which the organism possessed, even though the binary subjunctives did hold of it. (If alpha-centurions always have an intransitive structure of subjunctives hold true of them, is it *preferences* that they have?) An organism has preferences, it seems, not merely because and when binary subjunctives are true of it. The binary subjunctives which hold true of it must hang together in a reasonable fashion. The organism must show some modicum of rationality to be counted as having preferences at all. Given a being whose subjunctives usually and normally fail to rationally cohere (with no special explanation of this failure) it is doubtful that it is *preferences* which the being has.[39]

IV. TIME-PREFERENCE

Time-preference plays such a central role in the Austrian approach (to interest, capital theory, the business cycle, etc.) that we cannot close without giving it a brief examination. Böhm-Bawerk's famous three reasons for time-preference have frequently been examined[40], and we turn to more recent views.

    We can deal briefly with an argument of Rothbard's.

A fundamental and constant truth about human action is that *man prefers his end to be achieved in the shortest time*. Given the specific satisfaction, the sooner it arrives, the better. This results from the fact that time is always scarce, and a means to be economized. The sooner any end is attained, the better. Thus, with any *given end* to be attained, the shorter the period of action, i.e., production, the more preferable for the actor. *This is the universal fact of time preference* (Rothbard, *ibid., page 13, his italics.)*

This argument does *not* demonstrate time-preference on the basis of the scarcity of time. Time is scarce, and we want to economize it, to use less of it in achieving our goals. Given a way of achieving our goal that takes two minutes, and another way that takes five minutes, we will choose the two-minute way, thereby leaving the three extra minutes for the pursuit of some other goal. However, time-preference is not the same thing as economizing time. Suppose there are two acts A and B, which each take five minutes to do, and yield the same goal, but one delivers its goal earlier than the other. Suppose A delivers seven minutes after it's done, and B delivers one year after *it's* done. However, each taking five minutes to do economizes time

equally. Time-preference, therefore, cannot be derived from economizing time. It also may conflict with it, as when the action which delivers its goal sooner takes longer to perform than the action which delivers the goal later.

Mises connects time preference with action in a different and more striking way:

Time preference is a categorial requisite of human action. No mode of action can be thought of in which satisfaction within a nearer period of the future is not – other things being equal – preferred to that in a later period. The very act of gratifying a desire implies that gratification at the present instant is preferred to that at a later instant. He who consumes a nonperishable good instead of postponing consumption for an indefinite later moment thereby reveals a higher valuation of present satisfaction as compared with later satisfaction. If he were not to prefer satisfaction in a nearer period of the future to that in a remoter period, he would never consume and so satisfy wants. He would always accumulate; he would never consume and enjoy. He would not consume today, but he would not consume tomorrow either, as the morrow would confront him with the same alternative (*Human Action*, p. 484.)

On this view, an action shows time-preference because if the person didn't prefer doing it now, he wouldn't do it now. Furthermore, if the person never preferred doing it at a time rather than at a later time, he would never act at that time, and so would never act at all.

Let us scrutinize this argument more closely. First, a person might be indifferent between doing some act now and doing it later, and do it now. ("Why not do it now?")[41] So action now can show time-(weak) preference, but it need not show time-(strong) preference. Secondly, a person might act now to get a particular satisfaction, without caring whether it comes sooner or later. He acts now because the option of getting the satisfaction is a fleeting one which will not be available later. Thus, a person can have a reason, other than time-preference, to act now; to prefer satisfaction sooner rather than later is not necessary in order to act now. Thirdly, the fact that we act constantly cannot show that we *always* have time-preference for *all* goods. *At most*, it shows that when a person acts (and the option also is available later) he has time-preference *then* for the *particular* good he then acts to get. This is compatible with an alternation of periods of time-preference for good G, and periods of *no* time-preference for good G. The person acts to get G during one of the periods of time-preference for G. This is considerably weaker than general time-preference, as might be seen by considering what the theory of interest would look like if there were only this

weak form of time-preference. Finally, even if Mises' approach yielded the strong conclusion he envisages (which, I believe, it does not), it would leave time-preference mysterious. Action shows time-preference; but why is there time-preference? Time-preference would still stand in need of *explanation.*[42]

Let us approach time-preference by considering apparently unrelated propositions, which Austrian theorists usually list as additional (non *a priori*) assumptions, viz., leisure is a consumer's good, and labor has disutility.[43] It should be possible, I think, to derive the first of these from other deeper facts (which will not themselves be necessary truths, though). Note first that leisure *need not* be a consumer good, for each person. It depends what the person's *labor* is like. For example, suppose you have *no* desire for privacy, and a psychologist wants to study someone's life. He wishes to watch you from afar, unobtrusively, and he will pay you a fixed salary per hour. There is no limit to the amount of time you will gainfully labor, doing as thou wilt. Do not say that this is not *labor*, because it has no disutility, for that *would* make it a necessary truth that labor has disutility. Mustn't labor, at least, be an activity? The hired person is always acting. What he sells is a right; the right to watch him. It might even be more particular labor he does; his employment might require him *not* to do some trivial act *A* he cares nothing for.

Notice how the particular labor I have described differs from almost all other labor. It is a feature of almost all labor that it is incompatible with the simultaneous doing of other activities which the person wishes to do and which have some value for the person. Some consumption takes time, and cannot be done simultaneously with some other activities; for example, listening to Beethoven quartets and working a steam drill, lolling on the beach and teaching a philosophy class. Therefore, a person with multiple desires, some of which cannot be satisfied simultaneously with the particular labor he does, will want leisure time in which to satisfy these other desires. The value of an hour of leisure to the person will be the value to him of satisfying the most highly ranked of his desires or combinations of desires which can be satisfied in an hour, whose satisfaction is incompatible with the simultaneous doing of his particular form of labor. We do not need the additional assumption that labor has disutility. Even if labor is a good, we can still have tradeoffs of labor and leisure; that is, tradeoffs of labor with the other goods that can be had only in that person's nonlaboring time. (Thus, Austrian

writers need not deny that work may itself have intrinsic satisfying quality.)
These are empirical facts of a high level of generality:

(a) there are multiple goods;
(b) most people's labor (what others will pay them for) is incompatible
    with their simultaneously achieving some other goods (as they judge
    them). These other goods cannot be achieved (*as well*) simultaneously
    with the labor;
(c) in this area, beneficial tradeoffs are possible.

Now consider time-preference in the light of these.[44] If a person knew the
moment of his death, and knew each of the desires that he could satisfy, then
the incompatibility of various desires (the impossibility of their simultaneous
satisfaction) would require him, if all the desires are to be satisfied, to stack
desires back from the last moment, far enough so that they would all fit in.
This, however, will not constitute time-preference.

Suppose uncertainty is introduced; the uncertainty usually mentioned in
this context is uncertainty about the moment of one's death ("eat, drink, and
be merry for tomorrow we *might* die"), and about whether a currently
possible consumption will *continue* to be possible in the future. But it is
another uncertainty I wish to emphasize here (if only because it is generally
ignored), that concerning which *other* current desires (or which desires to be
acquired in the future), you may come to be in a position to satisfy in the
future. This uncertainty will lead a person to prefer satisfying some desire or
other *now* rather than postponing *all* their satisfactions until later. For it
might turn out later that another stronger desire can only be satisfied then,
which is incompatible with the satisfaction then of any of the desires which it
is currently possible to satisfy. And this might be true for all the 'laters.' (Add
to this the uncertainty about whether it will remain possible to satisfy one of
the current desires.) If you don't satisfy one of the desires now, you might
never do so, for later it might be outranked by an incompatible desire. In
these terms, there is nothing to *gain* by postponing the satisfaction of a
desire, and there is something which *might* be lost. Satisfying some desire
now, therefore, *weakly dominates* (in the sense of decision theory)
postponing the satisfaction of all desires until later. Hence we have the
rationality of an analogue of Mises' time-preference, though not of that which
involves a preference of greater current consumption and less future

consumption over uniform consumption (summing to the same as the previous alternative) throughout time.

The opportunity cost of failing to satisfy any desire now is the value of that desire, if any, current or future, which gets crowded out later. Since there might be such a crowded out desire, there might be this opportunity cost. Given two desires which can be satisfied now or later, the greater need not be satisfied first; the opportunity costs of satisfying it first might be larger than those of satisfying it second. The opportunity cost analysis explains the *sequencing* of the satisfaction of desires. There is no opportunity cost to satisfying a particular desire now rather than none, unless we imagine a reason to want it later rather than now. But starting, as we suppose, with no such preference either way, dominance considerations give us a time-preference.

But this does not appear to be the whole story about time-preference. Such preference has been discovered in *animal* experiments, where the effectiveness of a reward declines with its distance forward in time in accordance with some concave curve. If the curve is more concave than an exponential one, for example, if it is a hyperbola, then we have the possibility of two such curves *crossing*.
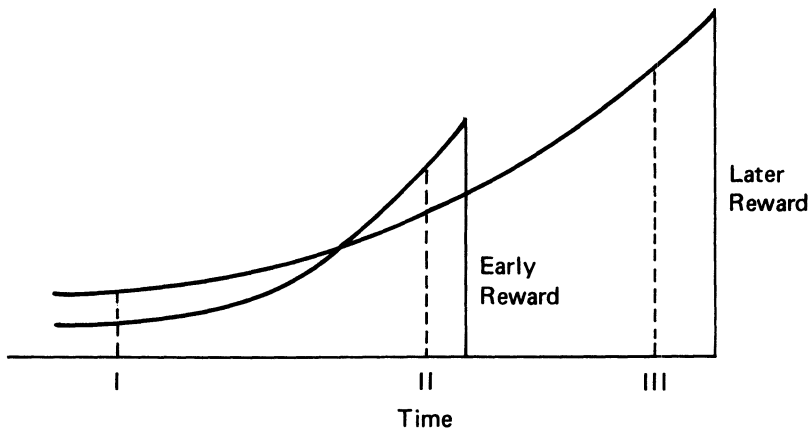
The height of a curve at each point in time represents how valuable the later reward is to the organism at *that* point in time. At each time an organism chooses to occupy the highest curve. Thus, in the diagram drawn, at time I a person would choose to wait for the larger later reward rather than take the

earlier smaller one which excludes getting the larger one, at time II the organism would choose the earlier smaller reward even though if it foregoes this and waits, the larger one will become available, and if the organism can somehow get beyond the time of the smaller reward, it will choose and reap the larger one. Thus, hyperbolic curves which cross are able to explain in an illuminating fashion the making and later breaking of resolutions (the first at time I and the second at time II), problems of self-control (at time II), etc.[45]

The time-preference found in animal experiments is not, I assume, to be explained by their performing rational calculations, even implicitly. How is it to be explained? If my earlier dominance argument about satisfying some desire or other in the present is correct, then an organism which does this and exhibits that sort of time-preference will, on the average, satisfy more of its desires. There would be other arguments as well which show that in a wide range of situations, time-preference would lead to greater total desire satisfaction. In an uncertain world, where contingencies frequently interfere with obtaining particular future satisfactions, time-preference enables organisms unable to reason about such contingencies to satisfy many of their desires.

Supposing such a time-preference tendency arose by random mutation and was transmitted genetically, and that the desires involved themselves were connected with survival to reproductive age, ability to protect progeny, etc., then time-preference would be evolutionarily adaptive, and would be selected for in the process of evolution, once it appeared. If some such explanation accounts for its presence in lower organisms, it is reasonable to think that we too have some genetically based time-preference. The evolutionary process has built time-preference into us, for within that process the rationality of time-preference is reflected as adaptive value.

My account here has been very sketchy, but some evolutionary account of time-preference in lower organisms must be given, and presumably we will be, in these ways, continuous with these other organisms. Economists should no more hesitate to use this biologically based fact about people than Mises hesitated to incorporate the general (non-*a priori*) statement that labor has disutility. However, it does leave us with a puzzle. The evolutionary process builds time-preference into organisms who do not calculate, as a (rough) rule of thumb to approximate what calculation would lead to. Discounting of future goods is a surrogate for the calculations of the dominance argument

(plus perhaps other maximizing arguments). Let us now consider the situation of organisms who do calculate, and who can in their deliberations take into account various future contingencies. If these organisms (call them people) do such explicit calculation, and *also* feed into these calculations magnitudes of (future) desires which have been discounted to take such calculative considerations into account already, then isn't there *double-counting*, or rather, *double-discounting*? Time-preference first discounts, and our later calculations explicitly take into account factors and lead, in effect, to explicit discounting. When we ourselves can take various contingencies explicitly into account, do we want to be bound by the type of discounting found adaptive as an average rule of thumb in the evolutionary history of the species? Shouldn't we try to correct for such discounting as is built into us, and put *all* discounting into our calculations? Or should we instead eliminate all such discounting from our *calculations*?

The problem of double discounting arises when our calculations utilize the current value to us of the future goods, which value we explicitly discount in order to take account of the uncertainties of the good's being realized. The discounting is *double* because the current value to us of the future good exhibits time-preference, an implicit discounting of the future selected for in the evolutionary process because of the adaptive advantage of discounting for future uncertainties and other such factors.

This problem of double discounting would be avoided if our calculations, instead of utilizing the current value to us of future goods, utilized some measure of how much the good would be worth to us at that time in the future when it would be realized. There would be no double counting in explicitly discounting such future values which have not yet been discounted even implicitly, for uncertainties.

However, a measure according to your preferences now of how valuable a future good will be to you *then* does not fit well with Austrian theory's close typing of preference to actual choices, or even to the loosening in terms of subjunctives which we suggested earlier.[46] So the problem of double-discounting is an especially pressing one for Austrian theory. Having built-in time-preference and also realizing what the evolutionary explanation of such time-preference is, how should we, as rational beings, behave?

Department of Philosophy,
Harvard University

NOTES

[1] This essay was written to accompany essays by economists in a volume of essays on Austrian economics. Since it was directed to an audience of non-philosophers, some of the philosophical points will not be news to readers of this journal. I have retained these points in the hope that some economists will find them illuminating, and that even philosophers might find their application interesting. I have benefitted from discussion of the essay at the New York Seminar on Austrian Economics, the Society for Ethical and Legal Philosophy, and from a discussion of Section IV with Richard Herrnstein.

The major writings in the Austrian tradition are Carl Menger, *Principles of Economics, Problems of Economics and Sociology*; Eugen von Böhm-Bawerk, *History and Critique of Interest Theories, The Positive Theory of Capital*; Ludwig von Mises, *The Theory of Money and Credit, Human Action*; Frederick Hayek, *Individualism and Economic Order, The Pure Theory of Capital, Prices and Production, Monetary Theory and the Trade Cycle*.

See also Ludwig von Mises, *Socialism*; Murray Rothbard, *Man, Economy and State*; Ludwig Lachman, *Capital and Its Structure; Macro-Economic Thinking and the Market Economy*; and Israel Kirzner, *The Economic Point of View, Market Theory and the Price System, An Essay on Capital*, and his recent and important *Competition and Entrepreneurship*.

Not every figure accepts or places equal emphasis on each of the features of the methodological framework. For example, while all parts of this essay apply to Mises (and to Rothbard) only Section I is clearly relevant to Hayek's views.

[2] See F. A. Hayek, *The Counterrevolution of Science* (The Free Press, Glencoe, Ill., 1955), Chs. 1–10, and *Individualism and Economic Order* (Univ. of Chicago Press, 1948), Chs. 2–4. For a sample of discussion in the journals see the six essays by Watkins, Mandelbaum, and Goldstein (two apiece) reprinted in Leonard Krimmerman (ed.), *The Nature and Scope of Social Science* (Appleton, Century, Crofts, N.Y., 1969).

A standard account of reduction is provided in Ernest Nagel, *The Structure of Science* (Harcourt Brace, and World, N.Y., 1961), Ch. 11. For complications and refinements, see also Lawrence Sklar, 'Types of Inter-Theoretic Reduction', *British Journal for the Philosophy of Science* 18 (1967), 109–124.

[3] I leave aside the question of how much of such decision theory as has been developed should be acceptable within Austrian theory, though I do believe that much of it is, and that mistaken views about the relationship of preference to choice and action (see Section III below) have led Austrian writers to ignore this literature. Indeed, one would expect that writings on personal or subjective probability would be most congenial to Austrians, the major proponents of the subjective theory of value. It is puzzling that Austrian writers have been concerned solely with 'objective' probability, since there is no guarantee that an actor will act on *those* ratios or limits of relative frequencies. Only if such ratios are reflected in subjective probabilities will a theory of human action be concerned with them, and if they are *not* so reflected, the theory will have anyway to be concerned with the (divergent) subjective probabilities. On personal (subjective) probability, see Leonard J. Savage, *The Foundations of Statistics* (John Wiley and Sons, N.Y., 1954) and F. J. Anscombe and R. J. Aumann, 'A Definition of Subjective Probability', *Annals of Mathematical Statistics* 34 (1963), 199–205.

⁴ I ignore here the issue of whether the *situation* can be specified by utilizing only the concepts of Crusoe theory.

⁵ See R. D. Luce and H. Raiffa, *Games and Decisions* (John Wiley and Sons, N.Y. 1957), Thomas Schelling's treatment of coordination games in *The Strategy of Conflict* (Oxford Univ. Press, 1963), John Harsanyi, 'A General Theory of Rational Behavior in Game Situations', *Econometrica* 34 (1966), 613–634, and 'Advances in Understanding Rational Behavior', Working Paper CP-366 of the Center for Research in Management Science, Univ. of California, Berkeley, July 1975, and the papers by Harsanyi cited therein.

⁶ See Luce and Raiffa, Ch. 9.

⁷ The non-methodological individualist might agree that the explanatory theory of n-person interactions suffices to explain all interactions, for he might believe that the ultimate explanation of n-person interactions lies in a general theory of multi-person $(> n)$ interactions. (I owe this preceding remark and the next paragraph to David Hills.) The methodological individualist, however, believes that the general theory of multi-person interaction won't make *essential* reference to more than n persons in that it will be non-vacuously true in some domain of no more than n individuals.

Furthermore, the theory where $n > 1$ will completely satisfy methodological individualist strictures only if it does not attribute utility or reasoning to dyads or groups in a way which is irreducible to individual utility functions and individual reasoning. It will not countenance an irreducible group mind, utility function, etc. (It is not obvious how to specify this 'etc.'.)

⁸ See Richard B. McKenzie and Gordon Tullock, *The New World of Economics* (Richard D. Irwin, Inc., Homewood, Ill. 1975), Ch. 18, 'Riots and Panic', for a methodological individualist account of such behavior.

⁹ On emergence, see Nagel, *ibid.*, pp. 366–380; Hempel, *Aspects of Scientific Explanation*, pp. 258–264.

¹⁰ Carl Menger, *Principles of Economics*, Ch. VIII.

¹¹ Mises holds that his money regression theory demonstrates that no other origin of money is possible. "It follows that an object cannot be used as money unless, at the moment when its use as money begins, it already possesses an objective exchange-value based on some other use. This provides both a refutation of those theories which derive the origin of money from a general agreement to impute fictitious value to things intrinsically valueless (here Mises footnotes Locke) and a confirmation of Menger's hypothesis concerning the origin of the use of money." (*The Theory of Money and Credit*, Enlarged edition, Yale University Press, 1953, p. 110) The prices tomorrow depend upon the amount spent on consumption, saved, and held in current cash balance. What is held in cash balance will depend upon expectations about prices tomorrow (the very ones to be explained) and prices the next day. It is not, strictly, circular to explain prices on a given day in terms of expectations about prices for that

very day, but one might well want to avoid this, especially if one contemplates the possibility that all such expectations are formed in accordance with (and *only* in accordance with) the very theory being developed. Mises stops the (unstrict) regress by having *today's* prices determining tomorrow's cash balances. Day by day we go back in time until the point when it is only because of use value, and not exchange value, that something is held. Thereby, all prices are non-circularly explained.

Mises is able to introduce today's prices as determining tomorrow's cash balances only if either

    (a) people expect tomorrow's prices to be (roughly) like today's

or

    (b) people arrive at their expectation of tomorrow's prices by starting with knowledge of today's prices, plus their knowledge of the causal factors operating, and thereby arrive at some view of how tomorrow's prices will differ from today's, and so of what tomorrow's prices will be.

In each case, it is an *expectation* about tomorrow's prices which is crucial to determining the cash holdings. Mises' money regression argument shows *one* way such expectations may function. However, it cannot demonstrate that Menger's account describes the *only* way money could arise, for it cannot demonstrate that expectations about tomorrow's prices can only be formed on the basis of today's prices. So, in particular, it cannot show that a social contract *could not* actually give rise to (roughly) uniform expectations about the next day's prices. To demonstrate *that*, another argument would be needed.

[12] Economics provides the prime arena of such explanations as have thus far been offered. George Homans presents illuminating examples of methodological individualist explanations in sociology in his *Social Behavior*, Revised edition (Harcourt, Brace, Jovanovich, Inc., N.Y., 1974), and explicitly endorses methodological individualism in *The Nature of Social Science* (Harcourt, Brace, 1967.)

[13] But things are more complicated if the explanation of the subjunctives and act-tokens essentially refers to previously holding subjunctives as well as to previous act tokens.

[14] Different theorists have held that capitalist institutions are self-destructive, notably Marx and Schumpeter, Most recently, Daniel Bell has advanced such claims in *The Cultural Contradictions of Capitalism* (Basic Books, N.Y., 1975).

[15] See E. O. Wilson, *Sociobiology*, Harvard University Press, 1975.

[16] Some writers speak not merely as if different psychological types inhabit the different historical stages of human society, but as if different psychological laws function in these different stages. I find this far fetched, but if it were true, and if the general law connecting the historical stages with the operative psychological laws was a brute-macro-law, unexplainable by any deeper psychological law as it operated in different social circumstances, then methodological individualism would be false. However, the explanations of behavior within each stage and of the transition of society

from one stage to another, might still be methodological individualist explanations which use the psychological laws of the relevant stage. What would not get a methodological individualist answer would be the question "why those laws?"

[17] For a recent survey of the current state of the theory, see Roger Brown and Richard Herrnstein, *Psychology* (Little, Brown, and Co., Boston, 1975), Chs. 1–3. I shall say something more about operant conditioning below.

[18] Mises himself holds that "leisure is a consumer's good" and "labor has disutility" are not necessary truths, and hence explanations which utilize these statements are *not* claimed to be reducible to a completely *a priori* theory.

[19] An illuminating survey of difficulties with the positivist arguments on synthetic necessary truths by a defender of such truths, is Arthur Pap's *Semantics and Necessary Truth* (Yale University Press, 1958). The most influential recent critic of necessity and of the analytic-synthetic distinction is W. V. Quine. See his essay 'Two Dogmas of Empiricism' in *From a Logical Point of View* (Harvard University Press, 1953), his book *Word and Object* (MIT Press, 1960), and his essay 'Necessary Truth' in his volume of essays *The Ways of Paradox* (Random House, N.Y., 1966). The most influential and ingenious recent defense and utilization of synthetic necessary truths, and of essences, is Saul Kripke's monograph, 'Naming and Necessity', in Donald Davidson and Gilbert Harman (eds.), *Semantics of Natural Language* (D. Reidel,|Humanities, N.Y., 1972).

[20] *Anarchy, State, and Utopia* (Basic Books, N.Y., 1974), pp. 42–45, 'The Experience Machine.'

[21] A survey is contained in B. F. Skinner, *Science and Human Behavior* (Macmillan Company, N.Y., 1953). A precise quantitative form of the law of effect is presented in R. J. Herrnstein, 'On the Law of Effect', *Journal for the Experimental Analysis of Behavior* 13 (1970), 243–266, and 'Quantitative Hedonism', *Journal of Psychiatric Research* 8 (1971), 399–412. It is often erroneously believed that Noam Chomsky's famous review essay of Skinner's book, *Verbal Behavior* (*Language* 35 (1959), 26–58, reprinted in J. J. Katz and Jerry Fodor, *The Structure of Language* (Prentice-Hall, 1964), refutes Skinner's general theory or at least its application to human beings. However, many of Chomsky's criticisms deal with earlier stimulus response theories, theories of drive reduction etc., and *not* with Skinner's theory of operant conditioning (which is not an S–R theory), and Chomsky's complaints about the circularity of the law of effect are mistaken. (See Paul Meehl, 'On the Circularity of the Law of Effect', *Psychological Bulletin,* 1950; Kenneth MacCorquodale, 'On Chomsky's Review of Skinner's Verbal Behavior,' *Journal of the Experimental Analysis of Behaviour* 13 (1970), 83–99; and the articles by Herrnstein cited above.) Even the claim that an act is self-reinforcing has the consequence (as Meehl pointed out) that successive performances will *increase* its strength and that artificial duplication of its proprioceptive or brain effects could be used to reinforce *other* actions. Nor is it empty to claim that exploratory behavior is itself reinforcing, for not only is this claim plausible on evolutionary grounds, but tests could determine whether being given opportunity to engage in exploratory behavior itself functions as reinforcement.

One of Chomsky's criticims has carried great weight with many, namely that (even if the law of effect itself is not circular and empty of content) many of the applications of the law of effect are circular. For in many applications, the reason for believing a reinforcing event has taken place is the very effect it is to explain. Consider another case. A bridge collapses, and the question of why it collapsed is raised. Engineers hypothesize that the stress in a particular place reached a certain point. They believe this occurred, because they have a theory, well confirmed elsewhere, that this causes bridges to collapse and is the most frequent cause of bridge collapses. They then go on to *explain* the collapse of the bridge by referring to the stress. The collapse of the bridge is their *reason for believing* there was a certain stress, and the stress is their *explanation* of the bridge's collapse. Isn't this circular? It is not, for the two italicized terms in the previous sentence are different. It *would* be circular if they were the same, that is if either (a) the collapse of the bridge is their *reason for believing* there was a certain stress, and the stress is their *reason for believing* the bridge did collapse, or (b) the collapse of the bridge is their *explanation of* the stress, and the stress is their *explanation of* the collapse of the bridge. But in the bridge example we are discussing, there is no such objectionable circularity. (See Jaegwon Kim, 'On Inference, Explanation and Prediction', *Journal of Philosophy* 61 (1964), 360–368. The situation is similar with reinforcement. Given a well-confirmed theory, it will be reasonable to infer some event of a type known to be reinforcing took place, or to infer that a type of event which was known to take place *is* reinforcing. One would want to go ahead and check this latter claim, but there seem to be no insuperable difficulties here. Thus, this criticism by Chomsky rests upon a mistaken methodological view. The theory of operant conditioning cannot be dismissed so easily.

[22] For a discussion of ways in which empirical procedures come into the discovery of necessary (though not *a priori*) truths, see Kripke, *ibid*.

[23] On verstehen, see F. A. Hayek, *The Counterrevolution of Science*.

[24] See Ernest Nagel, *The Structure of Science*, pp. 480–485.

[25] W. V. Quine, *Word and Object*, Ch. 2. Putnam's essay considers the relevance of *verstehen* for the issues Quine discusses about translation, and so links up the two areas, but it does not go on to treat radical verstehen and its implications for the empirical character of social science.

[26] Or, if we view empathetic understanding as assigning prior probabilities, after a large number of cases we will be able to assess its reliability as a prior probability assigner.

[27] This scheme is not identical, however, to Mises' *a priori* scheme. For most people do not think they *always* act to reduce their own felt uneasiness, etc. Hence, if Mises is right, these people's empathetic understanding of their *own* behavior is sometimes faulty.

[28] See Rothbard, *ibid*., p. 64.

[29] Mises writes in *Human Action*, "However, one must not forget that the scale of

values or wants manifests itself only in the reality of action. These scales have no
independent existence apart from the actual behavior of individuals." (p. 95); "The scale
of value manifests itself only in real acting; it can be discerned only from the observation
of real acting. It is therefore impermissible to contrast it with real acting and to use it as
a yardstick for the appraisal of real actions." (p. 102).

[30] Perhaps Mises is denying this point in the middle of the last paragraph of page 119 of
Human Action, but it is difficult to be sure. However, on p. 122 he says, "All
parts – units – of the available stock are considered as equally useful and valuable if the
problem of giving up one of them is raised." Here, then, we *do* have *indifference*. Yet a
choice will be made, perhaps at random. One particular object will be given up. Yet the
person does not prefer giving up *this one* to giving up another one. Therefore, choice
entails (at best) weak preference; it does not entail strong preference.
    Rothbard claims (*ibid.*, p. 265), "Any action demonstrates choice based on
preference; preference for one alternative over others. There is, therefore, no role for the
concept of indifference in economics or in any other praxeological science." However,
(on pp. 18–19) he also writes, "in these examples, the units of the good have been
interchangeable from the point of view of the actor. Thus, any concrete pound of butter
was evaluated in this case *perfectly equally* with any other pound of butter" (his italics
the first time, mine the second), and he continues on in the same vein on the rest of
p. 19.

[31] *Human Action*, p. 97.

[32] Can't we consider the cost of A as the value of choosing among the remaining
alternatives, making the implicit assumption that the value of choosing among all of the
remaining alternatives is equal to the value of the best alternative among them? Notice,
though, that this involves introducing a choice situation (choosing among *just those*)
which the person never actually faces. And making the implicit assumption explicit
would reraise the problem.

[33] See David Lewis, *Counterfactuals*, Harvard University Press. I am not committed
here, however, to the adequacy of Lewis' account, which in any case I describe only
roughly in the text. Furthermore, since the actual world is of course the world closest to
the actual world, Lewis' account has the unfortunate consequence of making
subjunctively true any conditional with true antecedent and consequent. We can avoid
this consequence by the following proposal, which has other desirable consequences as
well. Let the p-neighborhood of a world W be the closest band of worlds where p is true,
uninterrupted by a world in which not-p holds. More precisely, if p is true in world W,
the p-neighborhood of W is the set of those worlds w such that p is true in w and there is
no world w' in which not-p is true which is at least as close to W as w is. If p is false in
W, the p-neighborhood of W is the set of those worlds w such that p is true in w and
there is no world $w,'$ in which not-p is true which is the same distance from W as w is, and
for any world w'' which is between w and W and in which not-p is true, there is no world
in which p is true which is as close to W as w'' is. (The condition where p is false in W
says that the p-neighborhood of W is the set of those worlds w where p is true, such that
no not-p world is equidistant from W, and such that any not-p world between w and W is

in the not-p neighborhood of W.) However, when p is false, if the closest p-worlds to W are the same distance from W as are some not-p worlds, then a more complicated account of the relevant band must be given.

Finally, we can say that a subjunctive with antecedent p, and consequent q is true if q is true in each world of the p-neighborhood of the actual world.

[34] For an account of dispositions which views them as placeholders for an underlying physical structure see W. V. Quine, *The Roots of Reference* (Open Court, La Salle, Ill., 1973), pp. 8–12. See also Isaac Levi and Sidney Morgenbesser, 'Belief and Disposition', *American Philosophical Quarterly* 1 (1964), 1–12.

[35] The requirement that preference be transitive should *not* be read as, "if X is preferred to Y, and Y is preferred to Z, then X should be preferred to Z," so as to license an inference to the consequent if the antecedent is satisfied. Perhaps it's the case that X shouldn't be preferred to Y. Also, if the person prefers X to Y, Y to Z, and Z to X, (and so has intransitive preferences), three applications of the principle, read as above, would lead to the conclusion that he should prefer X to Z, Z to Y and Y to X – another intransitive triad! Instead, the requirement should be read as requiring preferences to hang together in a certain way, and as excluding preferences from combining in other ways, e.g., "it shouldn't be the case that X is preferred to Y, Y to Z, yet X isn't preferred to Z." See Robert Nozick, *The Normative Theory of Individual Choice* (unpublished doctoral dissertation, Princeton University, 1963), pp. 94 –98.

[36] Caution: see my remarks below about the connection of preferences with the subjunctives.

[37] For a recent survey of the theory of choice sets, the theory of conditions such as these, see A. K. Sen, 'Choice Functions and Revealed Preference', *Review of Economic Studies* 38 (1971), 307–317.

[38] As an antidote to Rothbard's statement (*ibid.*, p. 15), "It is important to realize that there is never any possibility of *measuring* increases or decreases in happiness or satisfaction . . . In order to any measurement to be possible, there must be an eternally fixed and objectively given unit with which other units may be compared," readers may wish to see D. Krantz, R. D. Luce, P. Suppes, and A. Tversky, *Foundations of Measurement*, Vol. 1 (Academic Press, New York, 1971).

[39] This type of point will be familiar to readers of W. V. Quine, and Donald Davidson.

[40] See, for example, Mark Blaug, *Economic Theory in Retrospect* (Irwin & Company), Revised Edition, pp. 503–509; Robert E. Kuenne, *Eugen von Böhm-Bawerk* (Columbia University Press, New York, 1971), pp. 25–34.

[41] This can happen only if the person at that moment is indifferent for all things he can get between getting them then and getting them later. Otherwise, he'll do one of the (other) time-preferred things.

[42] Those who think time-preference is a categorial component of action may think it strange to attempt, as I do later, an evolutionary explanation of it, and compare this to an evolutionary explanation of preference, another (supposed) categorial component of action. However, neither is inappropriate. Since it is not a necessary truth that there are preferences or time-preferences (or actions!), the world need not have contained them, their existence is a proper topic for explanation, and the correct explanation presumably will be some evolutionary one.

[43] See Mises, *ibid.*, pp. 131–132, Rothbard, *ibid.*, pp. 37–38.

[44] That consumption takes time, and may be incompatible with other activities, including other consumption, is put to different economic use by Staffan Linder, *The Harried Leisure Class*, Columbia University Press, 1970.

[45] I take these details from George Ainslie, 'Specious Reward: A Behavioral Theory of Impulsiveness and Impulse Control', *Psychological Bulletin* **82** (1975), 463–496, who also discusses in an illuminating way various devices for getting past the earlier smaller reward without choosing it.

[46] The relevant question is not what you actually would choose in the future if given the choice then (for your preferences might change), but how valuable, according to your preferences now, getting it then will be worth to you then. It is not clear how to capture this by a subjunctive about choices.