

# EXCERPT FROM *MORALS BY AGREEMENT*

DAVID GAUTHIER

## Chapter 2: Choice: Reason and Value

'There is nothing either good or bad, but thinking makes it so.'<sup>1</sup> But if things considered in themselves are neither good nor bad, if there is no realm of value existing independently of animate beings and their activities, then thought is not the activity that summons value into being. Hume reminds us, 'Reason is, and ought only to be the slave of the passions', and while Hume's dictum has been widely disputed, we shall defend it.<sup>2</sup> Desire, not thought, and volition, not 'cognition, are the springs of good and evil.

We might wish to bypass the great questions of value that have exercised Western philosophers from the time of Socrates and the Sophists. But answers to these questions are implicit in the theory of rational choice. Although the presentation of a full theory of value lies beyond the scope of our present enquiry, yet we shall find that in discussing the basic concepts of choice theory--preference, utility, maximization--we are led not only to outline a technical apparatus, but also to reflect on reason and value. In this chapter we shall consider both the apparatus and the assumptions needed to explain morality as arising from rational agreement.

But rational agreement itself, and its role in interaction, must wait to be introduced. Our present focus is on *parametric* choice, in which the actor takes his behaviour to be the sole variable in a fixed environment. In *parametric* choice the actor regards himself as the sole centre of action. Interaction involves *strategic* choice, in which the actor takes his behaviour to be but one variable among others, so that his choice must be responsive to his expectations of others' choices, while their choices are similarly responsive to their expectations.<sup>3</sup> Rational choice is well defined for parametric environments; we shall argue that morality enters in defining it for certain strategic environments. But this is to anticipate. How shall we understand parametric choice?

In Chapter I we spoke of the interest or interests of each person, relating rationality to the satisfaction of individual interest, and morality to an impartial constraint on the pursuit of that interest. But interest is a conception that hovers uneasily between an individual's own perspective and that of an outsider. The theory of rational choice takes as primary a conception even more clearly subjective and behavioural than interest, the relation of individual preference. Preference relates states of affairs; one speaks of preferring an apple to a pear, but more strictly one prefers the eating of an apple to the eating of a pear in some given environment or set of environments. The theory of rational choice is of course primarily concerned with preferences between states of affairs conceived as alternative possibilities realizable in action. These states of affairs are therefore not direct objects of choice, but rather are possible outcomes of the actions among which one chooses. The theory does not analyse particular relations of preference, which are treated as ultimate data, but sets of these relations, each set representing the preferences of one individual over the pairs of realizable outcomes in a choice situation.

Practical rationality in the most general sense is identified with maximization. Problems of rational choice are thus of a well-known mathematical type; one seeks to maximize some quantity subject to some constraint. The quantity to be maximized must be associated with preference; we have spoken loosely of advantage, or benefit, or satisfaction, but the theory of rational choice defines a precise measure of preference, *utility*, and identifies rationality with the maximization of utility. Utility is thus ascribed to states of affairs considered as objects of preference relations. The constraint under which utility is to be maximized is set by the possibilities of action. The rational actor maximizes her utility in choosing from a finite set of actions, which take as possible outcomes the members of a finite set of states of affairs.

An objector might agree to identify practical rationality with maximization, but insist that a measure of individual preference is not the appropriate quantity to maximize. It is rational to maximize value; the theory of rational choice implicitly identifies value with utility, but the objector challenges this identification. A radical challenger would insist that value is not a measure, but a norm or standard for preference. Value does not depend on

preference, but rather rational preference depends on value. We shall explain our rejection of this position in clarifying our conception of value. A more moderate challenger would agree that value is a measure, but insist that it does not measure brute preferences, which may be misinformed, inexperienced, or ill-considered. We shall accept this view in so far as it concerns the manner in which preferences are held. Thus we shall develop a set of conditions for *considered preference*, which must be satisfied if utility, as a measure of preference, is to be identified with value, and the maximization of utility with rationality.

We must also sketch the formal conditions for defining utility. We have said that it is a measure of preference; more precisely it is a measure of outcomes representing relations of preference. Given a set of these relations, constituting the preferences of one individual in a choice situation, then a utility is to be assigned to each possible outcome in such a way that one may infer the person's preference between any two outcomes from the utilities. For any two possible outcomes the one with greater utility must be preferred. But not all sets of preference relations permit the definition of such a measure. Suppose for example that Bruce prefers eating an apple to eating a pear, eating a pear to eating a peach, and eating a peach to eating an apple. Then the required measure of Bruce's preferences must assign a greater value to eating an apple than to eating a pear, to eating a pear than to eating a peach, and to eating a peach than to eating an apple. But this is impossible. Bruce's preferences violate the conditions on sets of preference relations which are necessary for the definition of a measure. Preferences failing to satisfy these conditions are incoherent; we must sketch the conditions of coherent preference.

To avoid misunderstanding, we should note that although we may infer preferences from utilities, we do not explain preferences by utilities. Utility is simply a measure; its role is similar to that of temperature. If the temperature in Aix-en-Provence is higher than the temperature in Toronto, we may infer that it is warmer in Aix than in Toronto. But we do not explain Aix's greater warmth in terms of its higher temperature. It is not warmer because its temperature is higher, but rather its temperature is higher because it is warmer. Measures do not in themselves serve explanatory roles.

Utility is assigned directly to possible outcomes, but choice concerns actions. If the outcome of each possible action may be known with certainty, then choice among actions is equivalent to choice among their outcomes. The maximizing problem to be solved in choosing rationally is then very simple; one selects that action yielding the outcome with greatest utility. (If there are several such actions, then one chooses indifferently among them.) A purely ordinal measure of the outcomes suffices; rational choice under certainty requires only that the outcomes be ranked. Since most of classical economics examines choice under certainty, ordinal utility meets the economist's needs.<sup>4</sup>

But in most real-life situations, even if the environment may be treated as fixed, its characteristics are not fully known, and so the outcomes of some choices are risky or uncertain. (In the theory of rational choice, risk is associated with objective probability, uncertainty with subjective.)<sup>5</sup> Choice among actions must then be based not solely on the utilities of their possible outcomes, but also on the probabilities of those outcomes. The maximizing problem is more complex and can in general be solved only given an interval measure of preference. From the utilities of the members of a set of possible outcomes we must be able to infer not only the preference ordering, but also the relative strengths of the preferences. If Bruce prefers eating an apple to eating a pear to eating a peach, then from his utilities we must be able to infer whether his preference for eating an apple over a pear is stronger than his preference for eating a pear over a peach.

The formal conditions on sets of preferences that are required to define utility thus fall into two groups--those sufficient for a merely ordinal measure, and those additionally required for an interval measure. After we have sketched these two groups of conditions we shall be able to extend our account of the solution of the maximization problem from the simple case of choice under certainty to choice under risk and uncertainty. Strategic choice, involving interaction, raises further issues deferred to Chapter III.

If the conditions for coherent preferences are met, we may introduce a measure of the objects of preference. If the conditions for considered preference are also met, we may identify this measure with value, and its maximization with rationality. We may therefore speak of the conditions for coherent and considered preference as conditions of rational preference. But we must not be misled in so speaking, for none of the conditions addresses the content of particular preferences. Not the particular preferences, but the manner in

which they are held, and their interrelations, are the concern of reason. Once more we find ourselves in agreement with Hume, in this case when he says that it is 'not contrary to reason to prefer the destruction of the whole world to the scratching of my finger'.<sup>6</sup> It may be contrary to reason to hold such a preference in an ill-considered manner, or to conjoin it with certain other preferences. But considered in itself we cannot assess its rationality. Of course we should not deny that such a preference may be symptomatic of severe emotional disturbance to the point of insanity. But its pathological character must be established by considering how it is acquired and maintained. Only confusion results from the all too common supposition that a malfunction of the affections must imply a failure of reason.

The conditions of rational preference do not require that the preferences of different persons be cast in a common mould. Each person's preferences determine her values quite independently of the values of others. Thus a state of affairs is characterized not by a single value, but by a set of values, one for each person into whose preferences it does or may enter. Value does not afford a single uniform measure of preference but a measure relative to each valuer. And although values are ascribed to states of affairs, the ascription is attitudinal, not observational, subjective, not objective. As a measure of preference value is and must be contingent on preferences for its very existence. Thus the theory of rational choice treats value as a subjective and relative measure, not as an objective and absolute standard.<sup>7</sup> These contrasts are controversial and demand commentary.

The theory of rational choice also treats practical reason as strictly instrumental. This is not quite implicit in the identification of rationality with maximization, for as we have noted, one might suppose that the quantity to be maximized was not a measure of but a standard for preference, a standard objectively inherent in states of affairs whose apprehension involves an exercise of reason. On such a view reason would not be merely instrumental, but would also be concerned with the ends of action. But in identifying rationality with the maximization of a measure of preference, the theory of rational choice disclaims all concern with the ends of action. Ends may be inferred from individual preferences; if the relationships among these preferences, and the manner in which they are held, satisfy the conditions of rational choice, then the theory accepts whatever ends they imply.

An instrumental conception of rationality is thus linked to the identification of value with utility. Were value a standard, then reason might have a role in determining this standard which would go beyond mere maximization. Were there neither standard nor measure, then reason would have no practical role whatsoever; there would be nothing to maximize. The theory of rational choice sets its course between the dogmatism of assuming a standard for preference and the scepticism of denying a measure of preference. Since our aim is to have rational choice take moral theory in tow, we shall want to be sure that this course avoids shipwreck.

We have raised three topics to occupy the remainder of this chapter. In the next section we shall discuss considered preference. Then we shall sketch coherent preference. And in the final section we shall defend the conception of value as a subjective, relative measure.

**2.1** To this point we have spoken as if preference may be distinguished from choice, not only conceptually but also operationally. We have supposed that a person's preferences over all pairs of possible outcomes are available as basic data, that a measure of these outcomes may then be defined from which her preferences may be inferred, and that a rule for maximizing this measure may then be introduced as a basis for choosing. Since preference and choice are treated as distinct, the rationality of choices may be assessed by determining whether they maximize preference fulfilment. But as the reader familiar with economics knows, this is not how the economist relates preference and choice, and consequently not how he understands utility and maximization. <sup>8</sup> For the economist choice alone is accessible to observation. Rationality is assumed wherever possible, so that a person is treated as behaving rationally if her choices may be given a maximizing interpretation. Value is then identified with whatever she maximizes. Preferences are not given as data but are inferred; a person's preferences may be identified with the members of any set of relationships between possible outcomes such that their measure is maximized in her choices. On the economist's view preference is revealed in choice and has no independent operational significance.

Whether or not this approach is adequate for economics, it does not suffice for moral theory. The economist supposes that a person behaves irrationally, or fails to maximize value, if and only if her behaviour cannot be

given a maximizing interpretation. Although reason and value play a limited normative role in prescribing and criticizing choice, that role is expressed in the single injunction, 'Maximize!' To say that one should maximize utility as a measure of preference adds nothing, since utility is simply identified with whatever one's behaviour may be interpreted as maximizing.<sup>9</sup>

If reason and value are to play normative roles in the framework for understanding human action afforded by the theory of rational choice, then it must be possible to give the injunction to maximize further content. If a person may behave in a maximizing way and yet not fully rationally, then maximization, although necessary, cannot be sufficient for rationality. But revealed preference is too impoverished a conception to provide this further content. If our only access to a person's preferences is through her choices, then whatever she maximizes must be the measure of those preferences. To move from the economist's account to a view of rational choice adequate to understand rational behaviour, we must begin by reconsidering the conception of preference.

Yet we must not suppose that the economist's focus on revealed preference is simply mistaken. Certainly choices reveal preferences; a conception of preference that treated the linkage between preference and choice as merely accidental would be even more evidently inadequate than one that makes preference parasitic on choice. What we must reject in the economist's account is the insistence that choice affords the sole access to preference. We must distinguish a behavioural dimension of preference revealed in choice, and an attitudinal dimension expressed in speech. Choosing an apple from a dish of apples and pears, one reveals a preference for eating an apple rather than a pear. Saying that one likes apples better than pears, one expresses a preference for eating an apple rather than a pear. Of course in some situation, choice is verbal; one reveals one's preference by expressing it. But in other contexts the two dimensions of preference are quite distinct and may be in conflict. Karen expresses a preference for reading philosophical works rather than watching situation comedies on television, but she spends her free time in front of her television set while her philosophical library gathers dust.

We are now in a position to add some content, although as yet by no means sufficient, to the injunction to maximize. Utility is a measure of preference, but not of either revealed or attitudinal preference taken in isolation from the other. It is a measure of both dimensions in so far as they agree. If the two are in conflict, then even if it is possible to define a measure for each, neither may be identified with utility. If a person's revealed and expressed preferences diverge, then her values are confused and she lacks an adequate basis for rational choice. A necessary condition of supposing that a person is maximizing value, and so acting rationally, is that it be possible to interpret her choices as maximizing the measure of her expressed preferences.

Lack of a measure common to behaviour and attitudes is a sign of irrationality. But we may not without further ado ascribe this irrationality to a person's choices because they fail to conform to her attitudes, or to her attitudes because they fail to express her choices. In specific circumstances one of these ascriptions may seem clearly justified. But in general we may conclude only that the divergence between choice and attitude indicates irrationality without locating it more precisely. For it would be overcome were either the person's choices to come to conform to what she expresses, or her attitudes to come to conform to what she chooses.

In many situations a person chooses without any verbal expression of preference. In other situations a person expresses a preference without being in a position to make a choice. Of course we do not suppose that in these situations the person fails to be rational, simply because it is not possible to measure both dimensions of her preference. Rather we stipulate that wherever possible, an individual would verbally confirm her choice, or behaviorally confirm her attitude. We assume the coincidence of revealed and expressed preference unless we have clear evidence of divergence. We assume a single measure of preference unless we are confronted with opposed measures. We accept the economist's assumption that rationality is to be ascribed to a person's behaviour in the absence of contrary evidence; we reject the extreme limitations on the evidence he allows.

**2.2** Choice may fail in several ways. Queen Gertrude chose the poisoned cup. A young lady says 'Yes' (whether to a proposal or to a proposition) and later, although perhaps only momentarily later, realizes her mistake. A tourist newly arrived in France orders a bottle of Nuits-Saint-Georges to accompany his sole

meuniere. We may say that Gertrude's choice was uninformed, that the young lady's was unconsidered, that the tourist's was inexperienced.

Gertrude's choice, although fatally mistaken, was not irrational. She had no reason to believe that the cup contained poison. She wanted drink and chose what she took to be drink. Had she had reason to suspect the cup, then, given that her aim was not suicide, her choice would have been irrational. But we need not attribute this to her preferences; her beliefs would have been at fault. Her choice would have reflected failure to assess correctly evidence concerning her circumstances.

It might be supposed that a person's values should be identified with the measure of her informed preferences. But this we reject. Instead we relate a person's preferences to her state of belief or information. Gertrude chose poison, but she did not believe what she chose to be poison, and her choice did not reveal a preference for poison. Characterizing her preferences in relation to her beliefs, we have no reason to fault them, or to refuse to identify their measure with her values.

In some cases we may be able to infer what a person would have preferred, had she been correctly or fully informed about her circumstances. Gertrude would have preferred a cup of wine. But in other cases we may be quite unable to determine what persons would have preferred, since they will have formed no preferences in relation to the actual facts. We may relate the Aztecs' preference for human sacrifice to their belief in the hunger of the gods for human blood, but how would their preferences have been altered had they held no such belief? We do not know and have no way of knowing. The preferences of the Aztecs were formed in the context of their beliefs; beyond that context they had, or need have had, no preferences. To suppose that each person must have a set of 'real' or 'true' preferences related to her actual circumstances, although often concealed by false or incomplete beliefs, seems psychologically implausible.

A full discussion of the role of belief in relation to choice and preference would take us too far from our present concern. We accept the general explanatory schema: choice maximizes preference fulfilment given belief. We reject the trivialization of this schema that results from denying independent evidential access to each of its terms—to choice, preference, and belief. Choice may go astray because of mistaken belief; choice may be irrational because of irrational belief formation. The identification of value with the measure of preference, and the rationality of preference-based choice, are not affected by these failings of belief.

From Gertrude we turn to the inexperienced tourist. We may hope that before his departure from France he will realize that he does not prefer red wine with fish. His initial choice reflects his unfamiliarity with French cuisine; he has no firm or clear preference. Perhaps his choice is irrational; he would have done better to consult the waiter or to peruse the food-and-drink section of his guidebook; but waiters intimidate and guidebooks overwhelm. It may be that, his choice reflects his preferences and beliefs as best he could ascertain them in the circumstances. But we should not want to take his initially revealed preference for red wine with fish as a basis for determining his values, since he lacks the experience to form a considered preference.

The young lady's acceptance of the proposal (or proposition) is equally not a basis for determining her values. She may not lack the experience to form a considered preference. But she fails to reflect before accepting. When she does come to reflect she realizes that her firm preference is expressed by 'No'. Does she change her initial preference, or does she correct it? Neither alternative is acceptable; the first would imply that her successive preferences stand on an equal footing; the second would imply that initially she misstates her preference. A better account is that without due consideration, she forms a tentative preference revealed in her acceptance, which she then revises. Any choice reveals (behavioural) preference, but not all preferences are equal in status, so that the rejection of a tentative preference is not to be equated with the alteration of a firm or considered preference. Of course not every rejection of a tentative preference need involve the formation of a firm one. The young lady's 'No' might have been followed by yet another 'Yes'. But we may suppose that in this case further reflection only confirms her second preference. And so we treat it as affording a basis for determining the young lady's values.

In the absence of full reflection on one's preferences choice tends to go astray. Again, this need not show it to be irrational. Circumstances may not allow adequate reflection; the gunman who says 'Your money or your life'

is unlikely to accede to my request for time to think it over. Or even if time permits, the investment of effort in reflection may be unproductive; faced with fourteen brands of laundry detergent on the supermarket shelves, I simply reach for the nearest box. But if I am deciding where to spend my vacation, I shall act irrationally if I am not satisfied that further reflection would leave my preference unaltered.

If one chooses without adequate experience of the possible outcomes, one may later wish that one could undo one's choice. If one chooses without sufficient reflection on the outcomes, one may wish later to alter one's choice, should this be feasible. In neither case would one suppose one's original choice to have been successful, or to have maximized the measure of one's considered preferences. If one supposes that nevertheless one chose rationally, one must consider that one's choice was based on the best estimate that one could make in the situation of one's considered preferences--those that would pass the tests of reflection and experience.

But considered preferences, although stable, need not be unchanging. Mark, aged fifteen, prefers listening to disco to listening to Bach; twenty years later Mark prefers listening to Bach. His preference for disco indicates, we may judge, a lack of musical sensitivity and discrimination. As he acquires this sensitivity his preference changes, and he comes to think his later preference superior to his earlier one. He comes to think it better to prefer listening to Bach. But does this not imply that when earlier he prefers listening to disco, his preference is ill-considered?

Mark's lack of musical sensitivity at fifteen may invalidate any musical judgements he makes, but not his musical preferences. That he will come to prefer to listen to Bach does not give him a real preference for Bach in conflict with a merely apparent preference for disco. His current preference for disco is neither unreflective nor inexperienced; listening to disco does not make him wish to undo his choice, and should he find himself compelled to listen to Bach, he is not led to wish that he had done so voluntarily. In later life he wishes that when younger he had listened more to Bach. But nothing in his earlier preferences would have made such a choice rational. Although one may regret past choices because of their failure to maximize the fulfilment of one's considered preferences, one may also come to regret them because the fully considered preferences whose fulfilment they maximized do not correspond to one's present preferences.

An objector may point out that our argument ignores Mark's judgement that it is better to prefer to listen to Bach. He regrets his past choices, not merely because his preferences have changed, but because he has come to consider his past preferences inferior. And choices based on inferior preferences are faulty. But in the sense in which this last claim is true, it is not applicable to Mark's past choices. To suppose one preference superior to another is simply to prefer the one to the other. If one is in a position to choose among one's preferences, then one rationally chooses those preferences one considers superior. To choose inferior preferences would be to contradict expressed preference by revealed preference. But Mark does not do this. At fifteen he does not consider a preference for listening to disco inferior, and in any event he is not concerned to choose among possible musical preferences.

One chooses rationally in endeavouring to maximize the fulfilment of those preferences that one holds in a considered way in the choice situation. That these may include preferences that one will come to prefer not to have had, is a troubling feature of human life, but not of rational choice. Equally as troubling for life is that one may have preferences that at the same time one would prefer not to have. Some persons prefer to smoke--and their preference may be fully considered and all too experienced--and yet they prefer not to smoke. It may be that smoking, although obviously fulfilling the preference to smoke, diminishes the likelihood that they will come to prefer not to smoke. These persons face a hard choice, but not one that raises problems for our account of rational choice. The strength of their preferences, and the effect that choice is likely to have on their fulfilment, determine what it is rational for them to do. But surely new objections must occur to the reader. Are we not bypassing questions of interests and of prudence? To these and related issues we must therefore turn.

**2.3** Preferences are considered if and only if there is no conflict between their behavioural and attitudinal dimensions and they are stable under experience and reflection. Value is the measure of considered preference, and rational choice involves the endeavour to maximize value. This view may be challenged by anyone of several related positions. One may prefer smoking another cigarette to doing without, but it may be

held irrational to choose to smoke since smoking is injurious to health and it is in one's interest to be healthy, whatever one may prefer. One may prefer pursuing one's lost love to seeking consolation elsewhere, but it may be held irrational to continue in a hopeless quest which only adds to one's misery rather than to gain greater happiness or satisfaction by directing one's efforts to some more attainable goal. One may prefer spending all one's earnings to gratify present desires rather than saving for one's old age, but it may be held irrational to maximize the measure of one's present preferences rather than prudently endeavouring to maximize the measure of one's considered preferences over the course of one's future life. 10

We cannot discuss exhaustively the alternatives that might be advanced to our position, but shall simply sketch our principal reasons for rejecting the three views just mentioned. In each case the seeming plausibility of the alternative depends on contrasting it with the economic view that choice is rational in maximizing revealed preference. Once attitudinal preference, or even more reflective preference is introduced into the basis of rational choice, the plausibility of the alternative vanishes.

A person who acknowledges an interest in not smoking exhibits irrationality if she simply continues to smoke, without either endeavouring to stop or expressing some stronger considered preference for continuing. But we need not explain this irrationality by contrasting interest with preference as a basis for choice. To acknowledge an interest in not smoking is to express a preference, other things being equal, for those states of affairs furthering the interest over those not furthering it. If this acknowledgement is ignored in choice behaviour, then there is a divergence between expressed and revealed preference.

But what of the person who does not acknowledge an interest in not smoking? The defender of the view that rational choice is interest-based will insist that this person nevertheless has an interest in not smoking, and chooses irrationally in ignoring it. But here there is no overt conflict between behavioural and attitudinal dimensions of preference. Are we then not committed to the view that the smoker's behaviour is rational?

Not necessarily. Her revealed preference for smoking may fail to be contradicted by an acknowledged contrary interest because of her failure to reflect sufficiently. This failure may be inadvertent, or it may be deliberate, since the person may know that were she to reflect, she would have to acknowledge an interest in not smoking. If we may suppose that her revealed preference would not survive reflection unopposed, then we may agree that her choice is not fully rational.

But it will now be asked, what of the person who on the fullest and most careful reflection and the most extensive experience does not acknowledge an interest in not smoking? What of the person who consistently refuses to take her health into any practical account whether in word or deed? An acknowledged interest is an expressed preference, but on our view a rejected interest has no standing in determining either a person's values or the rationality of her choices.

Here then we disagree with the defender of the rationality of interest-based choice. To take an unacknowledged and indeed rejected 'interest', rather than a clear, considered preference, as the ground of choice is in effect to treat the content of preference as subject to rational assessment. And this requires an objective conception of value. If we treat the proposal to ground choice in interest as a subjectivist alternative to our view, then it fails. For if it be agreed that values are subjective, then there is no ground for appeal beyond what a person acknowledges, given that she has reflected sufficiently and is fully experienced. Of course her beliefs may be irrationally held, but this will not impugn her values, although it may show them to be inapplicable or irrelevant in her true circumstances.

There are few persons who do not acknowledge health as an interest, and even fewer who would not acknowledge it given adequate experience and reflection. For all but those few it would indeed be irrational to behave as if health were of no account. Those who believe that smoking is inimical to health must, if they are rational, choose not to smoke, unless their interest in health is outweighed by some stronger and equally considered preference. We need not take interest as the basis of rational choice to reach this conclusion, and we find no plausibility to any stronger conclusion.

A person may of course come to acknowledge as in her interest what previously she rejected, and may then regret having ignored it. This acknowledgement may reflect a change in belief, or in preference, or it may arise

through the person's experience of ignoring the interest. In this last case the person's past choices have not been successful. Through inexperience she failed to ascertain correctly her considered preferences. Whether this failure indicates irrationality must of course depend on the particular circumstances.

The second position we must consider briefly is that rational choice is based on satisfaction or happiness, a measure of the quality of experience, or enjoyment or hedonic tone. No particular term suffices to express this proposed basis for choice, perhaps because it is unclear precisely what is to be measured. But what is essential is that there be a measure, distinct from utility and so independent of preference, applied to situations in virtue of the way in which they are experienced. The defender of happiness supposes that this measure is to be maximized in rational choice.

An immediate difficulty with this view is that there is in fact no measure of our experience whose credentials we may assess. But let us assume such a measure, and let what is measured be called *enjoyment*, which we shall treat in this discussion as a semi-technical term. Satisfaction or happiness is then defined as the measure of enjoyment, just as utility is the measure of preference, or temperature the measure of heat.

We suppose that people become acquainted with enjoyment and its measure, so that in forming their preferences they may take into account the expected satisfaction afforded by the possible outcomes of their choices. There are now two possibilities. The one is that their fully considered preferences do not focus exclusively on enjoyment, so that utility and satisfaction do not always correspond. Now this would surely be the best possible case against equating rationality with the maximization of satisfaction. If enjoyment is not acknowledged on full reflection to be the appropriate object of preference, then no argument can ever show that the maximization of its measure is the primary requirement of practical reason.

On the other hand it may be that fully considered preferences are exclusively concerned with enjoyment, so that the measure of expected enjoyment always corresponds to the measure of preference. Were this so, we should have made a profound psychological discovery about preference. We should have found it to have a single object—enjoyment. We should have found that it is possible to discover a person's real preference between any two states of affairs without appealing to either her choices or expressions of preference, without even requiring that she be acquainted with or even aware of the states of affairs, provided that we could determine how much enjoyment she would experience in each. And we should be able to explain a person's preferences in terms of enjoyment and its measure, so that we should be able to say that she preferred one outcome to another because the first afforded her greater satisfaction.

Were all this so, we should agree that satisfaction afforded the basis for rational choice. But this would not be opposed to our present position, for satisfaction would provide the basis for rational choice as the measure of the object of preference, enjoyment. But of course we have no reason to believe that preference has a unique, measurable object such as we have supposed enjoyment to be. It seems clear that preferences cannot be correlated with any single dimension or characteristic of the states of affairs that they relate. Indeed, it seems clear that preferences do not depend solely on the qualities of experience, as the defender of satisfaction or happiness maintains.<sup>11</sup> We may dismiss this alternative from further discussion.

The third position opposed to our preference-based account of rational choice identifies practical rationality with prudence. It is not rational, the defender of prudence claims, simply to maximize the fulfilment of one's present preferences, however considered they may be, if one does not take into account the preferences one will or may come to have. Of course one cannot have full knowledge of one's future, but one may form reasonable expectations, allowing for different alternatives with varying probabilities. The prudent person takes thought for the morrow and so, our objector claims, chooses rationally. An example will illustrate how prudence-based and preference-based accounts of rational choice differ. Winter will come, and when it comes Susan will prefer her home to be well insulated. Is it not rational for her to insulate now, when prices are lower and workers and materials readily available? This would be the prudent course of action, but it is summer, and Susan is reclining by her pool without any care for winter. She believes that she will care, but right now she does not. The defender of the rationality of prudence insists that her present unconcern is irrational. But on our account, only a consideration entering into one's present preferences can provide rational support for choice. If Susan's unconcern is fully considered, then she has no reason to insulate now.

Here we face the proverbial contrast between the grasshopper and the ant; are we not defending the rationality of the heedless grasshopper? Indeed we are, in so far as the grasshopper is reflectively heedless, but we are not thereby impugning the rationality of the prudent ant. To maximize on the basis of one's present preferences need not be to ignore one's future preferences; one may take an interest in one's future well-being now, preferring a satisfying life to more immediate gratification. But also, one may not. Our view is that prudence is rational for those who have a considered preference for being prudent, but not for those who on full reflection do not.

Our disagreement with the defender of prudence does not turn on whether future preferences are to be taken into account, but on how they are to be considered. We both agree that the unreflectively heedless person, who takes no thought about the morrow, chooses irrationally. One's considered preferences for possible outcomes of one's choices must take into account the expected effects of those outcomes on oneself. But one may choose to ignore those effects in what one does; one may choose to take no thought for the morrow. And this reflective heedlessness is not irrational on our view. The defender of prudence insists that rational choice must be directed to the maximal fulfilment of all our preferences, present and future, in so far as we are able to determine what they are. On our view rational choice must be directed to the maximal fulfilment of our present considered preferences, where consideration extends to all future effects in so far as we may now foresee them.

There is a close link between the defender of interest-based choice and the defender of prudence. Both view the future in a temporally neutral way, denying any priority to the present moment in rational deliberation about what to do. Our position rejects this temporal neutrality except in so far as it is part of an individual's present stance. Or in other words, we suppose that the extent of an individual's practical identification with her future selves is determined by her present preferences, and these preferences, as all others, are not subject to rational assessment in terms of their contents. To be sure, the markedly imprudent, devotees of instant gratification, may find that when they taste the bitter fruit of their choices, they regret their past folly. But their irrationality is a contingent matter, determined by their considered preferences, and not a matter determined conceptually by the alleged atemporality of reason. Practical reason takes its standpoint in the present. 12

Against the view that mere behavioural preference suffices as a basis for rational choice, and as a determinant of value, the defenders of interest, of enjoyment, and of prudence, may press plausible objections. But the view that rational choice rests on considered preference, exhibiting no divergence between behaviour and attitude, and manifesting adequate experience and reflection, assimilates the merits of these alternative positions. Considered preference is, however, but part of our story. Maximization itself imposes conditions on preference. But these conditions, required for the definition of an interval measure, are quite independent of the way in which preferences are held. Unconsidered preferences may meet, and considered preferences may fail, the conditions for maximization. To simplify our discussion we shall henceforth talk only of preference, leaving it understood that unless we should specifically state otherwise, our concern is with considered preference and with the maximization of utility taken as the measure of considered preference.

**3.1** The formal conditions required in order to define an interval measure of preference are extensively discussed in the literature of rational choice. 13 We have nothing to contribute to the mathematical treatment. Our aim in this section is merely to offer a relatively non-technical account of coherent preference. As we noted previously our discussion falls into two parts: first we sketch the conditions sufficient for a merely ordinal measure of preference; then we introduce the further conditions necessary for an interval measure. An ordinal measure may be defined over the members of a set on the basis of any relationship weakly ordering those members. The relationship places the members from first to last, best to worst, most to least, with ties permitted. Height weakly orders human beings from tallest to shortest. Lifetime major league batting average weakly orders baseball players from highest (Ty Cobb) to lowest (?). An ordinal measure of preference may be defined over the possible outcomes in any choice situation if and only if preference weakly orders those outcomes, from most preferred to least preferred (with ties--equal preference or indifference--permitted).

The first condition required if a relationship is to weakly order the members of a set is that any two members be comparable on the basis of the relationship. Thus for any two possible outcomes in a choice situation, the

chooser must either prefer one to the other or be indifferent between them. This is the condition of completeness; it rules out preferentially non-comparable outcomes.

As a condition on behavioural preference completeness is trivial. If a person must choose between two possible outcomes, then his choice will reveal a preference between them. But of course, he may be unwilling or unable to express or acknowledge a preference. As a condition on considered preference completeness is not trivial. Its failure does not preclude choice, but only rational choice. If someone finds two possible outcomes preferentially non-comparable, then he is unable to choose rationally between them. This is a direct consequence of the view that rational choice rests on preference.

But it should be noted that the need for a condition of completeness arises because rational choice is based on a relationship among the objects of choice. Whatever that relationship is, it must satisfy such a condition. If we were to substitute interest, or enjoyment, for preference, as the basis for rational choice, then we should have to be able to compare any two possible outcomes in terms of interest fulfilment or in terms of enjoyment, in order to choose rationally between them.

Much of our discussion in this section is concerned, not so much with preference itself, as with the requirements of any account identifying rationality with the maximization of the measure of some relationship among the possible outcomes of actions. But of course we must ask if these requirements may plausibly be imposed on preference; if not, we should seek some other relationship as the basis of rational choice—or we should give up the identification of rationality with maximization. We suppose that it is plausible to require that preference satisfy completeness, that it is plausible to claim that a person who is unable to compare two states of affairs preferentially is indeed unable to choose rationally between them.

The second condition required if a relationship is to weakly order the members of a set is that the relationship be transitive. Let us illustrate the transitivity condition as it applies to preference by a simple example. Suppose that among the possible outcomes of Bruce's choice are eating an apple, eating a pear, and eating a peach. If Bruce prefers eating an apple to eating a pear and eating a pear to eating a peach, then if transitivity holds he prefers eating an apple to eating a peach. If Bruce is indifferent between eating an apple and eating a pear, and eating a pear and eating a peach, then if transitivity holds he is indifferent between eating an apple and eating a peach. Furthermore, if Bruce prefers eating an apple to eating a pear and is indifferent between eating a pear and eating a peach, or is indifferent between eating an apple and eating a pear and prefers eating a pear to eating a peach, then if transitivity holds he prefers eating an apple to eating a peach.

Even more than completeness, transitivity represents a standard for preference rather than a characterization of actual behaviour and attitudes. In some contexts the standard is no doubt attained. If the possible outcomes are all sums of money then we may expect perfect transitivity of preference, a larger sum always being preferred to a smaller one and equal sums being treated as indifferent. But if there is no one quantitative dimension along which the outcomes may be ranged and which is the sole object of concern in choice among them, then transitivity may be expected to fail in the actual world. This is most evident in the case of indifference. One is indifferent between a first apple and a second, between the second and a third. . . . between the nineteenth and a twentieth, yet one is in no doubt about one's preference for the first to the twentieth. Each small difference between successive apples falls beneath one's threshold of discrimination, but summed the differences are clearly noticeable. Sufficiently similar objects or states of affairs are regarded as alike, but sufficient similarity is not a transitive relationship.

Gross failure to satisfy transitivity leads to a cycle of preference, as exemplified by our reference to Bruce in the first section of this chapter, who prefers eating an apple to a pear, a pear to a peach, and a peach to an apple. If manifest in behaviour, such cycles would divest choice of all rationality. Suppose that Bruce has an apple. Before he eats it we offer him a peach; he prefers the peach and so trades. Before he eats the peach we offer him a pear; he prefers the pear and so trades. But before he eats the pear we offer him his original apple; he prefers the apple and so trades. After three trades, in each of which Bruce betters his situation in terms of his preferences, he finds himself at his starting-point.

Indeed, Bruce's situation is even worse than this example suggests. If he prefers a peach to an apple, then presumably he is willing to offer more than the apple in return for the peach. Suppose he is willing to offer I

cent. And suppose he is willing to offer the peach and 1 cent for the pear, and the pear and 1 cent for the apple. Then after three trades, in each of which Bruce' betters his situation in terms of his preferences, Bruce finds himself with his original apple but 3c poorer. Clearly an astute trader can take advantage of Bruce's intransitive preferences to beggar him by an unending round of such trades. And clearly Bruce, who becomes ever less well off as a result of trades, each of which betters his position in terms of his preferences, is engaged in an irrational sequence of choices.

Acyclicity, or the avoidance of full preference cycles, must be required of any relationship if it is to ground rational choice. Full transitivity is a stronger requirement, demanded by the identification of practical rationality with maximization. But its plausibility does not rest on that identification alone. Were no more than acyclicity required, Bruce might be indifferent between eating an apple and eating a peach, even though he preferred eating an apple to a pear and a pear to a peach. In ruling out such preference combinations, transitivity surely expresses our pre-theoretic ideal of rationality in choice.

Completeness and transitivity are together sufficient to ensure that the preference relationship weakly orders any set of possible outcomes. Each such set then includes an outcome preferred to all others, or several outcomes, mutually indifferent, but each preferred to all others. These are the most preferred outcomes; utility, the measure of preference, is maximized by the selection of a most preferred outcome.

But as we have noted, the direct objects of choice are actions, not outcomes. Under conditions of certainty, in which each action is correlated with a definite outcome, this poses no problem for maximization. But under conditions of risk or uncertainty, in which each action is correlated only with a set of possible outcomes, each with its (objective or subjective) probability, it may not be possible straightforwardly to choose a most preferred outcome. Suppose Bruce prefers eating an apple to a pear to a peach, and that his preferences are transitive. There are two bowls of fruit, one with two apples, five pears and one peach, the other with four apples, one pear and three peaches. Bruce may choose a bowl, from which he will then receive one fruit selected at random. Which bowl should he choose? His preference ordering does not yield an answer, since it does not reveal the strength of his preference for apples to pears, in comparison with his preference for pears to peaches. If he very much prefers pears to peaches, but only slightly favours apples to pears, then the first bowl will be the better choice. But if he strongly prefers apples to pears, and only slightly favours pears to peaches, then the second bowl will be better. To determine the respective strengths of Bruce's preferences, an interval measure is needed.

**3.2** Under risk or uncertainty; each action open to a person's choice is associated with a set of outcomes. Each outcome in the set is assigned a probability; under risk, as we have noted, these probabilities are objectively based, whereas under uncertainty the probabilities are determined by the individual's subjective estimates. For present purposes the distinction between objective and subjective probability may be ignored, as long as we require that the subjective probabilities be assigned so that they sum to unity. It is taken as certain that if the action is chosen, one of the outcomes in its associated set will result, and only one.

We may think of each action as a lottery, with its associated outcomes as its prizes. But we should enter a warning about this terminology, which is standard in the literature of rational choice. Each actor's preferences for any choice treated as a lottery, must depend solely on his preferences for the outcomes or prizes and on their probabilities. The lottery itself is transparent to preference; each person is indifferent to the 'gamble' which the lottery involves. In real-life lotteries the participants are concerned, not only with the prizes and their probabilities, but with the pleasures (or displeasures) of taking a chance. But in the lotteries associated with risk and uncertainty, taking a chance is not one of the factors affecting preference.

Utility, the measure of preference, is determined by a real-valued function, that is, a function assigning real numbers to outcomes in such a way that, if outcome X is assigned a greater number than outcome Y, outcome X is preferred to Y, and if outcomes X and Y are assigned the same number, they are indifferent. Under certainty only preferences for outcomes enter into a person's choices, since each action is associated with a single outcome. But under uncertainty and risk, preferences for lotteries over outcomes must also enter into choice. If there is to be a measure of these preferences, such that we may interpret the person's choices as maximizing it, then two further conditions on preference, in addition to completeness and transitivity, must hold. Given these conditions, an interval measure over the outcomes and lotteries possible in a particular

choice situation may be defined, such that the utility assigned to any lottery is its expected utility--that is, the sum of the products obtained by multiplying the utility of each of the prizes (or outcomes) of the lottery by its probability.

An example may be helpful. Suppose Bruce is to receive a fruit taken at random from the first of the bowls mentioned at the end of the last subsection--the bowl with two apples, five pears, and one peach. Then the probability of Bruce receiving an apple is 0.25, a pear 0.625, and a peach 0.125. If the utility, to Bruce, of an apple is 1, a pear 0.5, and a peach 0, then the expected utility of receiving a fruit taken at random from the bowl is  $[(1 \times 0.25) + (0.5 \times 0.625) + (0 \times 0.125)]$ , or 0.5625.

To say that the utility of a lottery is its expected utility, is to say that preferences over lotteries may be inferred from the expected utilities of the lotteries. If lottery X has greater expected utility than lottery Y, then lottery X is preferred to lottery Y. Given utility as an interval measure of Bruce's preferences over possible outcomes, we may then infer his preferences over lotteries with these outcomes as prizes by calculating their expected utilities.

The theory of rational choice identifies practical rationality with utility maximization. If the utility of a lottery is its expected utility, then in conditions of risk or uncertainty practical rationality is identified with expected utility-maximization. This is the central thesis of Bayesian decision theory. <sup>14</sup> Although the theory does not command universal assent, we shall not enter into its merits and defects here, but only assert (dogmatically) that we know no satisfactory alternative.

What are the further conditions on preference required for an interval measure? The first is monotonicity. Consider two lotteries, X and Y, that differ in only one respect--one of the prizes (or outcomes) in X, P, is replaced by another prize in Y, Q (with, of course, the same probability). Monotonicity requires that if Q is preferred to P, then Y is preferred to X.

Consider a simple example. Diana is to receive a fruit taken at random from a fruit bowl. Initially the bowl contains three apples, three pears, and four peaches. But before Diana gets her fruit, Mark removes one of the peaches and replaces it by an apple. If Diana prefers getting an apple to getting a peach, then monotonicity requires that she prefer the final lottery to the initial one.

The second condition is continuity. Consider three possible outcomes, X, Y, and Z, such that X is preferred to Y and Y to Z. Continuity requires that there is one and only one lottery, with X and Z as prizes, that is indifferent to Y. Or in other words, there is a unique probability  $p$  such that the lottery  $[pX + (1-p)Z]$  is indifferent to Y.

Given that an individual's preferences satisfy completeness, transitivity, monotonicity, and continuity, there is an easy way (in principle) to assign utilities measuring his preferences over the members of any set of possible outcomes, such that the utility of any lottery taking members of the set as prizes will be its expected utility. Let the most preferred outcome(s) be assigned a utility of 1, and let X be such an outcome. Let the least preferred outcome(s) be assigned a utility of 0, and let Z be such an outcome. Then the utility of any other outcome Y is  $p$ , where Y is indifferent to the lottery  $[pX + (1-p)Z]$ . This assignment, it should perhaps be noted, is not unique; if U is a utility function providing an interval measure of an individual's preferences over the members of a set of possible outcomes, then any positive linear transform of U is also such a function. But we need not pursue these complications here.

We have not commented on the plausibility of monotonicity and continuity. Monotonicity in effect rules out preferences directed at lotteries themselves--preferences for or against risks--as opposed to preferences for the prizes (possible outcomes). We indicated at the beginning of this subsection that this must be done; the reader may, however, object that in many situations the requirement is unreasonable. If I prefer sunshine tomorrow to rain, then I must prefer a slight chance of sun (and otherwise rain) to rain for sure, if monotonicity is to hold. But surely, and quite reasonably, I may not; a slight chance of sun may only complicate my plans for the morrow in comparison with the certainty of rain. My preferences may be perfectly coherent even though I prefer sun for sure to rain for sure and rain for sure to a 10 per cent chance of sun.

Continuity may seem equally suspect. To borrow an example from one of the standard sources (Luce and Raiffa): I prefer \$1 to 1 cent and 1 cent to death. If continuity holds, then there must be a lottery over \$1 and death which I consider indifferent to 1 cent. But would I run any increased risk of death, in return for \$1 if I do not die? If not, then I must prefer 1 cent to any lottery over \$1 and death.<sup>15</sup>

There are seemingly easy responses to the objections revealed in these examples. The first, it may be noted, reflects misunderstanding of monotonicity and of the special features of the lotteries to which it applies. Suppose that I am indifferent between a holiday next year in Greece and a holiday next year in Mexico. If I am rational, must I be equally indifferent between either holiday for sure and a lottery with these as prizes? If the lottery is to be decided now, then it is plausible to hold that the answer is affirmative. But if the lottery is to be decided the day before the holiday commences, then the answer may well be negative; I do not care whether I take a holiday in Greece or Mexico but I want to know now where I shall go, so that I may make suitable plans. In this second case, the lottery introduces a period of epistemic risk that would otherwise be absent, and towards which I need not be indifferent.

Monotonicity entails indifference to risk or uncertainty as such. But it does not require indifference between the presence and the absence of a temporal period of uncertainty. It is a plausible requirement for coherent preference over lotteries in which the actual prize or outcome is determined immediately, so that the risk associated with the lottery does not affect the epistemic state of the participants. But real-life lotteries need not be decided immediately. Although I prefer sun tomorrow to rain tomorrow, I may prefer the certainty of rain to a small chance of sun because the weather 'lottery' will not be decided until tomorrow, leaving me in the subjectively undesirable state of not knowing what to expect. Such real-life lotteries do not undermine monotonicity as a condition on coherent preference, but they do reveal the limits of its applicability.

The second example, suggesting the failure of continuity, is perhaps best handled by limiting the contexts in which continuity is expected to hold by exempting extreme cases. But is there any precise way of distinguishing the extreme? Are we appealing to anything other than a gut response--usually continuity holds but sometimes our preferences turn out to be hierarchically or lexicographically ordered? With these questions unanswered, and further problems about monotonicity, amply represented in the literature of utility theory, unraised, we turn from our sketch of coherent preference, aware that the theory of rational choice is still an underdeveloped territory, to equally intractable but less technical concerns. <sup>16</sup>

**4.1** The introduction of utility as a measure of preference is in itself an intellectual exercise that does not determine the role utility is to play. We have claimed that by relating it to considered preference, we could treat it as the norm for choice, identifying rational choice with the maximization of utility. This is to equate utility with value. Economists have been content to note this and to proceed with their study of decision-making. Philosophers must pause, to examine the conception of value implicit in equating it with utility,

Value on this conception is a measure of preference. A measure depends for its existence on what it measures--no preference, no value. Of course value may be ascribed to an object--an apple, for example--or more properly to a state of affairs involving an object--eating an apple--without supposing that the object or the related state of affairs does in fact enter into the preferences of any person. The apple may be sufficiently similar to others actually involved in human preferences to permit an inference to the hypothetical preferences of some person concerning states of affairs in which it might be involved. And in this way value may be assigned to those states of affairs. But objects or states of affairs may be ascribed value only in so far as, directly or indirectly, they may be considered as entering into relations of preference. Value is then not an inherent characteristic of things or states of affairs, not something existing as part of the ontological furniture of the universe in a manner quite independent of persons and their activities. Rather, value is created or determined through preference. Values are products of our affections.

To conceive of value as dependent on affective relationships is to conceive of value as *subjective*. The opposed, objective conception does not deny that values are affectively related, but it does deny that they are products of such relationships. To conceive of value as objective is to conceive of it as existing independently of the affections of sentient beings, and as providing a norm or standard to govern their affections. The subjectivist view denies the existence of such a norm.

Here one might ask whether a theological conception of value, treating it as dependent on the divine will, is objective or subjective. Dependence on God's will, it may be urged, is the analogue to dependence on sentient affection. We may reply that on most theological conceptions, everything is dependent on the divine will, so that from the standpoint of finite, sentient creatures, value has the same objectivity as other aspects of the divine order. To conceive value as a divinely ordained norm or standard to govern our affections is to conceive it as objective.

The difference between subjective and objective conceptions of value is not a difference between the objects or states of affairs to which value may be ascribed. It might be thought that subjectivism implies that only what is itself subjective--only a state of sentient experience--could possess intrinsic value. But subjectivism concerns the ground of value, not its object. There is no restriction on the nature of those states of affairs that may be objects of preference, and so that may be valued. There may be some persons who reveal and express preferences only in relation to their own experience, but there are surely few whose preferences are so confined. We must not confuse the view that only what is preferred for its own sake is or can be intrinsically good, with the view that only what is experienced pleasurably can be intrinsically good.

An objective conception of value is equally neutral with respect to the nature of the objects of intrinsic value. It is quite compatible with, although certainly it does not require, the view that the only bearers of intrinsic value are states of sentient experience. One might suppose that contemplative activity was the sole, or more plausibly the supreme, intrinsically valuable state of affairs, and while one might claim that it was also most truly pleasurable and most appropriately preferred, these judgements of its affective character would derive from rather than determine its value, which on this view would depend on the cognitive nature of the experience.

Or consider a more mundane view--that individual conscious survival is the sole intrinsic good. A subjectivist might advance this position because he supposed that conscious beings capable of choice did in fact take conscious survival as their sole goal, at least given full consideration. An objectivist might advance this position even though he supposed that some conscious beings did not make survival their sole concern, but would insist that nevertheless it afforded the proper objective for rational choice. The subjectivist would base the value of conscious survival on its actual place in our concerns; the objectivist would determine its proper place on the basis of its value. But both would take sentient experience as affording the sole intrinsic value. If both are mistaken (as we should argue), their mistake is not one that offers any ground for choosing between subjectivist and objectivist conceptions of value.

Subjectivism is not to be confused with the view that values are arbitrary. As we have insisted, if utility is to be identified with value, it must be the measure of considered preference. Values are not mere labels to be affixed randomly or capriciously to states of affairs, but rather are registers of our fully considered attitudes to those states of affairs given our beliefs about them. Although the relation between belief and attitude is not itself open to rational assessment, it is not therefore arbitrary. We have noted our agreement with Hume's insistence that it is not contrary to reason to prefer the destruction of the world to the scratching of one's finger. A person who on full consideration held such a preference, would correctly assign to the destruction of the world a value greater than he would assign to the scratching of his finger. Although, as we have also noted, we might have grounds, independent of the content of this preference, for holding such a person to be mad, we should not have grounds for considering his preference arbitrary. Madness need not imply either a failure of reason or a failure of reflection. (Madness, we should hold, is primarily a disorder of the affections. But this does not imply that the affections are irrational, any more than a disorder of the stomach implies that it is irrational.)

Subjectivism is also not to be confused with the view that values are unknowable. Evaluation, as an activity of measurement, is cognitive. Preference, what is measured, is knowable. What the subjectivist denies is that there is a knowledge of value that is not ordinary empirical knowledge, a knowledge of a special realm of the valuable, apprehended through some form of intuition differing from sense-experience. Knowledge of value concerns only the realm of the affects; evaluation is cognitive but there is no unique 'value-oriented' cognition.

Value is the norm for choice. The objectivist considers value to be the norm for choice because it is also the norm for preference. For him, relating choice to preference does not show choice to be rational except in so far

as preference itself may be related to value as its norm. The subjectivist considers value to be the norm for choice because it is the measure of considered preference. For him, relating choice to preference does show choice to be rational, provided preference is fully considered and is coherent, fixed in relation to reflection and experience and satisfying the conditions necessary to define an interval measure. For the objectivist preference mediates choice and value; for the subjectivist value mediates choice and preference.

**4.2** Value is a measure of individual preference, on the conception we are defending. The value of a state of affairs for some person measures its place in his preferences. Each state of affairs is thus characterized, as we noted in the first section of this chapter, not by a single value, but by a set of values, one for each affective relationship into which it enters or may be thought of as entering. The relationship between the values of different persons, or between the different values ascribed to a particular state of affairs, is a matter for empirical investigation. It may be thought that divergent values reflect opposed beliefs, not all of which can be correct. Yet it may also be thought that convergent values reflect the uncritical acceptance by one individual or the members of one group of the ideas of others, to the exclusion of full individual consideration. We have no good reason to suppose that the fully considered preferences of different persons over the same states of affairs would give rise to equivalent measures, even if we suppose the persons fully informed, fully reflective and experienced, and similarly placed with respect to the states of affairs.

To conceive of value as dependent on each individual's own affective relationships is to conceive of value as relative. There are of course other relativist conceptions of value. Common to all forms of purely individualistic relativism is the view that each person has his own good (and bad), and that the goods of different persons are not parts of a single, overall good. Other, non-individualistic forms of relativism may suppose there to be a good common to a family, or to a class, or to a society, while denying that the goods of different families or classes or societies are parts of an overall good. Opposed to all forms of relativism is an absolute or universal conception of value. An absolutist conception holds that values are the same for all persons, or for all sentient beings. An absolutist may suppose that values are affectively determined, so that a state of affairs is good simply because it is preferred by some person, but then he must hold that the state of affairs is good, not only in relation to or from the standpoint of that person, but from the standpoint of every person. The individualistic relativist supposes that your good and my good are quite independent, and not parts of a single universal good; the absolutist insists that your good is simply that part of the single good which pertains to you. 18 If we were developing a full theory of value we should have to examine the nuances of both relativist and absolutist positions. We might find, not two opposed views, but a continuum of outlooks. But we shall remain with a simple opposition between a purely individualistic relativism and an undifferentiated absolutism.

On both the absolute and the relative conceptions it is possible to distinguish what is good for some person from what is straightforwardly good, and to suppose that a state of affairs that is good for some individual may be considered good from the standpoint of others. A state of affairs is good for someone if it contributes to his well-being. On the absolute view, if contributing to someone's well-being is a source of positive value, then it is such a source for everyone; a state of affairs that is good for-and perhaps only for a single individual is then good from everyone's standpoint in being good for that individual. On the relative view, contributing to someone's well-being is a source of positive value for others if it enters into their good, as it does on our particular relativistic position if those others prefer such a contribution. A state of affairs that is good for a single individual is then good from the standpoint of each person who prefers or would prefer the enhancement of that individual's well-being.

Perhaps the classic philosophic formulation of a conception of value both subjective and relative was offered in the seventeenth century by Thomas Hobbes. Writing in *Leviathan*, he says,

But whatsoever is the object of any mans Appetite or Desire; that is it, which he for his part calleth Good: And the object of his Hate, and Aversion, Evill . . . For these words of Good, Evill . . . are ever used with relation to the person that useth them: There being nothing simply and absolutely so; nor any common Rule of Good and Evill, to be taken from the nature of the objects themselves; but from the Person of the man.<sup>19</sup>

And again he says,

Good, and Evill, are names that signifie our Appetites, and Aversions; which in different tempers, customes, and doctrines of men, are different: And divers men, differ not onely in their Judgement, on the sense of what is pleasant, and unpleasant to the tast, smell, hearing, touch, and sight; but also of what is conformable, or disagreeable to Reason, in the actions of common life. Nay, the same man, in divers times, differs from himselfe; and one time praiseth, that is, calleth Good, what another time he dispraiseth, and calleth Evill. 20

Hobbes links subjectivism with relativism--the view that value is dependent on appetite or preference with the view that value is relative to each individual. Natural as this linkage may seem, so that the two positions are easily confused, yet each may be held without the other. Indeed, a common, although surely mistaken, interpretation of Hobbes himself denies that he is a subjectivist, while granting that he is a relativist.<sup>21</sup> Hobbes supposes that the chief good from the standpoint of any person is his own survival. As we have noted, this may be given an objective interpretation--individual survival provides a norm for our affections and preferences, so that anyone who seeks his own destruction is irrational, even if on full consideration he prefers death to life. Were this Hobbes's view it would be difficult to relate it to the position expressed in the passages we have quoted, in which value is clearly linked to appetite in a subjectivist manner.

A position both subjectivist and absolutist seems implicit in the views of many defenders of one of the most influential modern moral theories, utilitarianism. John Stuart Mill suggests such a position in his attempt to offer a sort of proof for the principle of utility--subjectivist in saying that 'the sole evidence it is possible to produce that anything is desirable is that people do actually desire it', and absolutist in insisting 'that each person's happiness is a good to that person, and the general happiness, therefore, a good to the aggregate of all persons' <sup>22</sup> But there is an evident awkwardness in this union of subjectivism and absolutism noticeable in Mill's own statement, which in passing from a seemingly relativist premiss (that each person's happiness is a good to that person) to an absolutist conclusion (that the general happiness is a good to all persons) has generally been held to exemplify the fallacy of composition.

Utilitarianism finds itself under pressure to move away from a conception of value at once subjective and absolute. The most plausible way to resist this pressure would seem to be to accept a universalistic conception of rationality, and to argue that since rationality is identified with the maximization of value, and rationality is universal, then what is maximized, value, must similarly be universal--the same from every standpoint. If however utilitarianism remains true to its roots in the economic conception of rationality, then either subjectivism or absolutism gives way. On the one hand value may be conceived as relative, but a special form of value, moral value, is introduced, which is the measure of those considered preferences held from a standpoint specially constrained to ensure impartiality. On the other hand value may be conceived as objective, as the measure of an inherent characteristic of states of experience-enjoyment-that affords a standard or norm for preference. This is not the place to embark on a discussion of these positions, so that we shall merely (but dogmatically) affirm that a hundred years of ever more sophisticated efforts to avoid Mill's fallacy have not advanced the cause of utilitarianism a single centimetre. But we shall of course give more serious attention, especially to the second of the above ways of defending utilitarianism, as we continue the exposition of our own theory.

If there is pressure against a conception of value at once subjective and absolute, there is also pressure against a conception both objective and relative. The view that each individual is a member of some natural kind, and that each kind has its own characteristic perfection, quite different from that of other natural kinds, is, if not widespread in secular ethics today, yet of great historical importance. But this view has rarely, if ever, embraced a relative conception of value. For the objectivity of each characteristic perfection, its role as a norm or standard against which each individual member of the kind may be judged, has been supposed to depend on considering each perfection to be a manifestation, appropriate to its particular circumstances, of a single universal good. The seemingly relative goods of the several kinds are really facets of absolute good. The demands of objectivity thus force an apparently relative conception of value into an absolutist mould.

We shall therefore treat a relative conception of value as the natural counterpart of a subjective conception, even though the two are logically independent. But relativism may seem an absurd doctrine. A McIntosh is a good eating apple (indeed, I am enjoying one as I write these words). Were you to deny this you would be mistaken. Yet you may be no apple fancier, or your taste in apples may extend to the sweet Golden Delicious

rather than to the tarter McIntosh. In calling a McIntosh a good eating apple the word 'good' is not 'ever used with relation to the person that useth' it; indeed, one would be wrong to claim that a McIntosh was a good eating apple if one's sole ground were one's own liking for it. Hence it would seem that the evaluation is not a relative one. Yet it is subjective; the goodness of McIntosh apples is independent of the taste of any particular individual, but not of taste in general. The value of a McIntosh is a measure of preference, but of common preference.

Plausible as this argument against relativism may seem, it exploits a confusion, present in Hobbes's thought and not uncommon in formulations of a relativist position, to advance a thoroughly mistaken objection. Hobbes is wrong to suppose that such words as 'good' are used in relation to the user; indeed, our typical evaluative terms presuppose a common standpoint which mayor may not be shared by the evaluator. A good eating apple is thus one that is commonly or usually preferred for eating. But although committed to this standpoint by the language of evaluation, one is not committed to it either as a measure of one's own preferences or as a standard for one's own choices. In saying that a McIntosh is a good eating apple, you do not imply that you are an apple fancier, or that you enjoy McIntoshes more than most other apples, or that you would or should choose a McIntosh to eat. A McIntosh is good, relative to the standpoint presupposed in the use of 'good', but this value does not afford a norm for choice from any other standpoint.

The utility of a state of affairs is its value, to a person, as a possible outcome--as an alternative, therefore, to certain other possible states of affairs. Beginning from these utility values, subjective and relative as they are, we may move in two quite different directions. We may generalize from the utilities that measure a particular person's preferences over different states of affairs, to the kind of values held by that person. From these personal values we are not able directly to infer particular preferences, but rather the kinds of preference that the person will exhibit in different situations. We may assess these personal values in the ways discussed in previous sections of this chapter, determining how far they are considered and coherent.

But we may also generalize from the utilities that measure the preferences of particular individuals to the range of values exhibited by the different states of affairs into which a particular object, or kind of object, may enter. We find that in some cases the preferences of different persons proceed from shared or similar standpoints; this is revealed by parallel orderings in their utilities. The common language of value exploits these parallel orderings; we may talk of good eating apples because we find similarities in the tastes of those who eat apples.

A common language of value may however be established in a more systematic way without requiring parallel tastes. For we may conceive the objects, or at least some of the objects, that enter into the directly valued outcomes, as related to each other in a market. They are commodities, items of exchange among persons. So conceived, each object comes to possess a fixed value in exchange, relating it to all other commodities, and common to all of the persons whose exchange relationships are embraced by the market.

The standpoint for evaluation induced by a market is not only recognized by each person in his evaluations, but is actually shared by him. Parallel preferences among apple fanciers determine a common standpoint from which a McIntosh is evaluated as a good eating apple; you may recognize this evaluation without sharing the preference for McIntoshes. But if a McIntosh has greater market value than a Golden Delicious, if it commands 25c and a Golden Delicious only 20c, then, whatever your tastes, you would prefer a McIntosh, since you may exchange it for a Golden Delicious and 5c. To prefer a Golden Delicious to a McIntosh within the context of exchange would be to prefer 20c to 25c. But the common values and common standpoint of the market rest on individual utilities which are strictly relative. In understanding how the absolute or universal values of the market emerge from a given set of relative individual values and a given technology of production, we in fact understand the fallacy in the charge that value relativism ignores the common standards underlying ordinary evaluation. Since the theory of value in exchange is a commonplace of elementary economics we shall not expound it here, but it is to that theory that the reader should turn if she finds herself tempted by the most common objections to a relativist conception of value. 4.3 The defence of subjectivism is the primary task facing anyone who identifies value with the measure of considered preference. The defence of relativism is essentially a matter of detailed enquiry into the relation between individual and common norms. But the conflict between subjective and objective accounts of value lies at the very core of moral philosophy. No disagreement more seriously hinders the development of an adequate moral theory. We cannot expect to resolve this conflict

in a way that all will find convincing or even plausible. Nevertheless we find the case for subjectivism compelling, and we shall endeavour to sketch our reasons-reasons which, we should note at the outset, owe much to the work of Gilbert Harman and John Mackie.<sup>23</sup> Our first problem arises in formulating the opposed, objectivist position. We face the inevitable danger of erecting a straw man. In particular we must beware of ascribing to the objectivist ontological commitments that she believes unnecessary or avoidable, and then defending subjectivism by rejecting this ontology. Let us say that the objectivist holds that any adequate account of our experience or our environment must refer to value or to the valuable as being independent of sentient beings and their affections. In our view this does commit the objectivist to an ontology of value but we shall not press this point in our argument.

We must immediately clarify what is implicit in the independence of value ascribed to objectivism. The objectivist may insist that there is a necessary linkage between sentient beings and value, holding that value provides the norm for our affections. The proper object of preference is, and is necessarily, the good. But this link, on the objectivist view, is not found within sentient affection, but ties that affection to something else which in itself affords the ground of the tie. This other terminal of the link (and we shall not impose on the objectivist a particular account of what it is) must be independent of sentient beings and their affections, even though linked to them. In our discussion, we shall designate this alleged other terminal, objective value.

How are we to decide whether any adequate account of our experience or environment must refer to objective value? The answer, in our view, is that we consider whether any reference to objective value occurs in the best explanation we can provide for our actions and choices.<sup>24</sup> Now we cannot embark here on a theory of explanation. So we shall only assert that the only serious candidate for an explanatory schema for human action is, as we noted previously: choice maximizes preference fulfilment given belief. Objective value plays no role in this account. The best explanation we can provide for our observations is that there are physical objects with properties that, given our sensory apparatus, cause those observations. We therefore accept physical properties as part of any adequate account of our experience and environment. But the best explanation we can provide for our choices is not that there are objects with values that, given our affective apparatus, cause those choices. We therefore reject objective values as part of any adequate account of our experience and environment. The difference between G. E. Moore's notorious simple properties, yellow and good, is that from the standpoint of explanation good is a fifth wheel.<sup>25</sup> Objective value, like phlogiston, is an unnecessary part of our explanatory apparatus, and as such is to be shaved from the face of the universe by Ockham's razor.

John Mackie has argued that the best case against objective value turns on its ontological queerness.<sup>26</sup> It does not fit comfortably with our account of apprehension, or of knowledge; it does not relate happily to 'fact'. The objectivist may complain that Mackie is imposing ontological and epistemological requirements that are no part of her position. But objective value would be, if not queer, yet unique in the kind of force or power that it would exert. It must exercise a form of attraction as the proper object of preference that is very different from any other attraction we observe or experience. (Yet is this so? Might not the attraction of objective value resemble that exercised by certain persons? 'From the moment she entered the room, every man had eyes only for her. You could not say that her appearance, or her manner, or her attire, was in itself striking, but no one could ignore her.')

However, we should not regard queerness as a disqualification for objective value, given the particles and forces to which our best physical explanations refer. What is queer about objective value is that it has no explanatory role to play, and that does disqualify it.

Mackie suggests that 'if the requisite theological doctrine could be defended', then a form of value objectivity would be true. <sup>27</sup> This seems correct. For the defence of such a theological doctrine would be, or would include, a defence of its explanatory role. And if we relate value to divine will, then in such a doctrine objective value would also play an explanatory role. We cannot deny that in principle, the best explanation for our actions and choices might require reference to objective value; indeed, our principal intellectual ancestors--the classical Greeks and medieval Europeans--advanced explanations with just this requirement. But our explanations are not their explanations. When Laplace doomed God to explanatory redundancy, stating that 'Je n'avais pas besoin de cet hypothese', he implicitly doomed objective value as well. <sup>28</sup>

There is an evident objection to our claim that objective value has no explanatory role. Value-ascriptions frequently enter into the reasons offered for our choices and actions. 'Why did you send your son to Harvard?'

'Because it's a good university.' This is not a very illuminating explanation, but it is at least the beginning of an explanation, whereas 'Because I prefer Harvard' is not.

Part of the explanatory role of value-ascriptions may be understood on the basis of our discussion of the common language of value in the preceding subsection. We relate our choices and actions to what is commonly preferred, or to what is preferred in exchange. In so far as the criteria for such preference are generally known, then it is possible to infer from value-ascriptions to particular characteristics; a good carving knife is sharp and keeps its edge. The explanatory role of such value-ascriptions is readily accounted for on a subjective and relative conception of value.

We admit that value-ascriptions are thought to have a further explanatory role. For the belief in objective value is widely, if confusedly held, and those who believe in objective value appeal to it in accounting for their behaviour, just as those who believe in God appeal to Him in accounting for their behaviour. But although we must refer to belief in objective value in the explanation we provide for certain choices and actions, we are clearly not thereby committed to such belief. That we cannot understand what religious persons do without referring to their belief in God does not commit us to such a belief. (That we cannot explain what certain would-be geometers do without referring to their belief that it is possible to square the circle does not commit us to such a belief.) Among the beliefs that enter into the explanation of our choices are some that, were they true, would require references to objective value. What we deny is that such beliefs are true.

Our account of value must therefore include what Mackie calls an 'error theory'.<sup>29</sup> We suppose that persons objectify their preferences, and so come to consider their subjective attitudes and affections as properties characterizing the objects of their preferences. If we were to suppose that the correct conception of value could be discovered by an analysis of ordinary language, we should no doubt be led to an objective conception (or perhaps to a conception with objective and subjective elements intertwined in hopeless confusion). But if instead we suppose that the correct conception of value can be discovered only by an appeal to the best explanation of what value is supposed to affect, then we uncover the error present in ordinary views, and establish a subjective conception. To be sure, inference to the best explanation will not show that there is no objective value, any more than it will show that there are no fairies at the bottom of the garden. We are content to put objective value on a par with the fairies.

Value, then, we take to be a measure of individual preference--subjective because it is a measure of preference and relative because it is a measure of individual preference. What is good is good ultimately because it is preferred, and it is good from the standpoint of those and only those who prefer it. We have tried to fill in some of the qualifications and niceties for this view to be convincing, but we repeat that a full development and defence would require a full study--not merely our brief sketch. Our concern is to demonstrate the possibility and the characteristics of a rational morality, given that value is itself subjective and relative.

## NOTES

1 Shakespeare, *Hamlet*, II. ii. 255-7.

2 Hume, *Treatise*, ii. iii. iii, p. 415.

3 For the distinction between parametric and strategic rationality, see J. Elster, *Ulysses and the Sirens: Studies in rationality and irrationality* (Cambridge, 1979). pp. 18-19, 117-23.

4. See J. C. Harsanyi, 'Advances in Understanding Rational Behavior', in *Essays on Ethics. Social Behavior, and Scientific Explanation* (Dordrecht, 1976), p. 94

5. See Harsanyi, *Rational Behavior and Bargaining Equilibrium in Games and Social Situations* (Cambridge, 1977), p, 9.
6. Hume, *Treatise*, ii. iii. iii, p. 416.
7. There are significant differences in the uses of 'subjective' and 'objective', 'relative' and 'absolute'. We shall seek to make our use of these terms clear in sect 4. 1-2 infra.
- 8 See R. D. Luce and H. Raiffa, *Games and Decisions* (New York, 1957), pp. 50-1, and D. M. Winch, *Analytical Welfare Economics* (Harmondsworth, Middx., 1971), p.25.
9. ... "We assume that individuals attempt to maximize utility, and define utility as that. .... the individual attempts to maximize.' Winch, p. 25.
10. Our concern is with ideal types, rather than with particular positions in the literature. A useful discussion based on interest, which rejects an objective conception of interest but collapses interest into satisfaction, is found in L. W. Sumner, *Abortion and Moral Theory* (Princeton, 1981), pp. 162-86. Note that our own reference to interest in ch. I was intended as a stand-in for preference and utility, and not for the specific notion discussed here. An influential discussion of prudence is found in T. Nagel, *The Possibility of Altruism* (Oxford, 1970), pp. 27-76.
11. Sumner states, 'If something enhances my welfare, it must, directly or indirectly, immediately or in the long run, either produce states of mind I find agreeable or prevent states of mind I find disagreeable', p. 184. So if value is the measure of individual welfare, or of happiness or satisfaction, an egoistic conception of value is established by conceptual fiat.
- 12 'The correct perspective on one's life is from now', B. Williams, 'Persons, character and morality', in *Moral Luck* (Cambridge, 1981), p. 13. See, in addition to Williams's paper, D. Parfit, 'Later Selves and Moral Principles', in A. Montefiore (ed.), *Philosophy and Personal Relations* (London, 1973), pp. 137-69. See also the discussion of the position that 'the notion of rational man should be extended so as to include temporal considerations', in Elster, pp. 65-77.
13. See Luce and Raiffa, ch. 2, and Harsanyi, *Rational Behavior*, ch. 3.
14. See Harsanyi, *Rational Behavior*, p, 9.
15. Luce and Raiffa, p. 27.
16. Elster continues the passage quoted in note 14, ch. I, above, 'but many problems concerning its scope and resolving power remain to be settled'. He then offers 'a list of such problems'. It is not a short list.
17. My colleague Shelly Kagan did ask just this, and numerous other questions leading to discussions and clarifications at many other points.
18. See G. E. Moore, *Principia Ethica* (Cambridge, 1903), pp. 97-102, for a defence of this form of absolutism.
19. Hobbes, *Leviathan*, ch. 6, p. 24.
20. Ibid., ch. 15, p. 79.
21. See B. Gert's introduction to Hobbes, *Man and Citizen* (New York, 1972),
22. John Stuart Mill, *Utilitarianism* (London, 1863), ch. iv, para. 3.
23. See G. Harman, *The Nature of Morality: An Introduction to Ethics* (New York, 1977), ch. 1; also J. L. Mackie, *Ethics: Inventing Right and Wrong* Harmondsworth, Middx., 1977) ch. 1.

24. Our approach is suggested by Harman, *The Nature of Morality*, pp. 6-9. The phrase 'the best explanation' comes from Harman, 'The Inference to the Best Explanation', *Philosophical Review* 74 (1965), pp. 88-95.
25. See Moore, pp. 7-10, for his comparison of yellow and good.
26. See Mackie, pp. 38--42.
27. *Ibid.*, p. 48.
28. See E. T. Bell, *Men of Mathematics*, 2 vols. (Harmondsworth, Middx., 1953), vol. i, p. 198.
29. See Mackie, pp. 35, 48-9.