The Good and the True

Ronald B. De Sousa

Mind, New Series, Volume 83, Issue 332 (Oct., 1974), 534-551.

The Good and the True

RONALD B. DE SOUSA

My target in this paper is a certain dogma about the rationality of wants. The dogma is that the formal conditions for the consistency of wants mirror the formal conditions for the consistency of beliefs. In so far as the relation of wanting to the good is analogous to the relation of believing to the true, this amounts to the doctrine that the 'structure of the good' is analogous to the 'structure of the true'. The good, like the true, can never conflict with itself: hence it is as irrational to want contraries as to believe them. I shall refer to this doctrine as 'the Unity of the Good'. A few impressionistic historical reminders should suffice to show that it has behind it much of our philosophical tradition.

In Plato's middle Dialogues, all good things have a single cause and exemplar in the Form of the Good. The connection of the Unity of the good with the unity of the true is attested by Plato's penchant for the doctrine that the *virtues* are all one, and that the one which they all are, is knowledge.

In Aristotle's *Nicomachean Ethics*, some doubts are expressed about the unity of the Form of the Good, on the ground that 'good' can be predicated in all the categories. One might try to infer from this that Aristotle thought two goods might conflict, where they belong to different categories: but I do not see how to work this out. In any case, Aristotle appears to settle for an architectonic view of the organization of ends, which is another version of the Unity of the Good. Conflicts are seen to be merely apparent as soon as one takes account of the subordination of all ends to the comprehensive political good [*NE* 1094b.]. Perhaps as a consequence, Aristotle has bequeathed us a smooth way of getting around the most striking class of counterexamples to the unity of the good: I refer to his explanation of Tragedy as springing from the hero's 'tragic flaw'. (A more plausible view of Tragedy is Hegel's, who sees it as a 'collision' of right with right.)

In Aristotle, however, there is no logical link between the Unity of the Good and epistemology: though it happens that the Political Good has something to do with securing the peaceful pursuit of

knowledge by the comfortable magnanimates, it is the unity of this end that is relevant to the Unity of the Good, not the fact that it consists in the pursuit of knowledge.

In Kant, on the other hand, the link is explicit. The Rational Will wills only what might become universal law of nature: and of course this means that the will is irrational if it wills incompatible states of affairs, precisely because such states could not be consistently believed in.

Among substantive ethical theories currently in the field, forms of Utilitarianism are dominant. Ignoring mutations that have transformed the theory, I shall consider two ways in which classical Utilitarianism is committed to some version of the Unity of the Good. The first way is that the Principle of Utility is supposed to be the single standard by which all judgments of value are made. Any apparent conflict is accounted for much in the manner of Aristotle: either it results from a failure to attend to the hierarchic relations of the principles in question to the Primary Principle—in which case the conflict is simply settled by appealing directly to the Primary Principle—or else two incompatible courses of action are found to have the same expected utility, in which case they are strictly *indifferent*: and indifference is not conflict. Mill himself had an honorable loss of nerve on this point. Or at least that is how I think we should view his Quixotic thrashings on the subject of 'qualitative differences' of 'higher and lower pleasures'. His hunch here seems to have been that certain pairs of pleasures— Pushpin and Pushkin—are strictly *incomparable*. Nevertheless, he thought we must compare them anyway, and inevitably he urged for comparisons of quality the method already pre-empted for comparisons of quantity: namely actual preferences, under the right conditions.

This brings me to the second way that Utilitarianism is led to the Unity of the Good. The doctrine is derived—fallaciously, as I shall argue—from an insistence on Behaviourism. The argument goes something like this: the only empirical significance of the concept of utility lies in what value people actually set on available alternatives; and the only empirical access to those values are actual choices made. (Remember Mill's famous equation of the desirable with the desired.) But in any situation that presents an exclusive and exhaustive alternative, only two outcomes are possible: A is chosen over B, or B is chosen over A. It is allowed that the choice may on some occasions be a random one (one that does

not reflect an asymmetric relation of preference between the two alternatives) and thus we have a total of three possible relations. A is preferred, B is preferred, or A and B are indifferent. This is the assumption of *connexity*: any two alternatives are comparable. An individual's preference ranking determines a *utility scale* unique for that individual up to linear transformation. (This simply means that a person's utility measure will have an arbitrary scale and an arbitrary zero point, but fixed *ratios* of all *differences* of value.) Thus all values are measured on a single scale—at least for a given individual.[1]

It is easy to see that this is just another version of the Unity of the Good. For on this view there can be no genuine conflicts. Where there is no dominance of one course of action over another, the alternatives are simply indifferent: they cannot conflict, since they are in point of value exactly equivalent. Compare the role of money for the pure financier: between two incompatible money-making schemes, the expected dollar yield is either different, in which case the choice is clear, or equivalent, in which case either will do. A handy way of persuading oneself on this is to ask what there could be to *regret* if alternative A is chosen over B, where A and B have the same expected utility. The answer cannot be merely: alternative B: for what makes B desirable? Its utility. And how much of it is there? n utiles. But that is nothing to regret, for you got n utiles, from A. And if you claim that there was something about B that you miss anyway, then either you are abandoning the utilitarian scheme by denying that all value can be tallied up on a single scale (for there are values not on that scale, which you now miss), or else you are cheating, counting in utilities (which you now miss) that you neglected to feed into the original computation.

Note that on this scheme it is not, of course, irrational to want two different and incompatible things, if 'want' is taken loosely. There are innumerable wants that are never to be satisfied, simply because something else is wanted *more*. This is an important point to bear in mind for the sequel. Nevertheless, in one sense of wanting, 'over-all-wanting' or 'wanting-all-things-considered', it is irrational on this view to want A rather than B and also want B rather than A. Decision theorists put this by saying that such a pair of wants violates the axioms of rational preference. It is in this

---

1   I gather that until recently economists have tended to shun interpersonal comparisons of utility. But the trend may be changing. Cf. Richard C. Jeffrey, 'On Interpersonal Utility Theory', *Journal of Philosophy*, lxviii, 20 (1971), 647–656.

strong sense that *wanting* is to be understood in the context of this discussion.

In this version of the doctrine of the Unity of the Good, the connection with the unity of the true is less obvious. But one might try to make it out thus. Since the behaviourism on which the view rests attributes wants on the basis of actual choice, it naturally brings in the kind of consistency that facts are subject to: for actual choices must result in actual states of affairs. So one is easily led to believe that wants must be such that their simultaneous satisfaction constitutes a possibly actual state of affairs.

This inference from truth functional consistency to the Unity of the Good is a *non-sequitur*. For what is wrong with inconsistent belief sets is that they provide a guarantee of falsehood. That there is something wrong with believing what is false is self-evident (and correspondingly hard to argue). So the pursuit of consistency is part of the avoidance of falsehood. No such justification is available for the demand that wants be consistent. Nothing is self-evidently wrong with wanting what is not the case. The notion that it is wrong seems to rest on the assumption that rationality should be defined by reference to an ideal world. We might formulate it thus:

(1) *If one's wants and beliefs are such as to guarantee by virtue of the relations between them alone (and not by reference to any actual facts) that one is not in an ideal world, then one is not rational.*

In an ideal world, all beliefs are knowledge and all wants are satisfied. So any inconsistency in your beliefs is irrational, since it precludes your beliefs from all being true; and any inconsistency in your wants is likewise irrational since it precludes their simultaneous satisfaction. Note how close we are to Kant, in this Utilitarian company. The ideal world is the Kingdom of Ends.

This conception of rationality is one for which I can find no argument (except that it is neat). Since it is all too evident that what is good need not be true, it is simply gratuitous to assume that what is thought to be good must be *possibly* true. I shall refer to this assumption as *Pejoratism*, since it implies that the world is worse than it might be. *Optimism*, the view that this world is the best one possible, is actually a rather gloomy view: for together with the facts of life it implies that no good world is possible. I am defending a variant of Gloomy Optimism: the view that any possible world lacks some good.

Before proceeding in this defence I need to characterize a little

more abstractly and clearly the view I am attacking. As I have expressed it, it is first that 'the formal conditions of consistency of wants mirror the formal conditions of consistency of belief'. Take that to mean that

(2) for any set S of propositions $[p_1, \ldots p_n \ldots]$ it would be inconsistent to *want* (each member of) S iff it would be inconsistent to believe (each member of) S.

It follows from this, for example that it is irrational to want p and q where 'p' implies '$\sim q$'.

The second formulation was that 'the good can never conflict with itself'. This is to be taken to mean that

(3) any set S of propositions $[p_1 \ldots p_n \ldots]$ is a set of *goods only if* S is consistent. (Cf. any set S is a set of truths *only if* it is consistent.)

(2) and (3) together form a partial characterization of the parallel I am attacking. As will become clear below (p. 542), however, there are also other disanalogies which must be noted.

In stating the view in these ways, it should be obvious that I am making certain presuppositions about the objects of wants and beliefs and the bearers of truth and goodness. I want to make two of these presuppositions explicit and say a word in their defence; I shall state a puzzle in connection with each, and this roundabout way will lead me back to my main topic.

The first presupposition is that *good* is the formal object of wanting, as *true* is the formal object of believing. This rather arch phrase means roughly what the Medievals expressed by saying: we want things *sub specie boni* and believe them *sub specie veri*. A formal object, given in answer to the question, 'why do you want [believe] X?' is vacuous but perfectly appropriate: 'because it is good [true]'. That is why it is *formal*: anything placed in the object slot of a belief or want expression must be or purport to be *a good* or *a truth*. (Counterexamples, which people often like to make up, are uncontroversially instances of irrationality.) Note that there is something odd about the 'because' in the above sentences: it is not causal, nor is it logical. Perhaps it is teleological: truth and good are the *targets* of belief and want. This metaphor suggests another way of indicating the relevant sense of *'formal'*: 'One shoots at the target' is a formal remark (a remark about the 'grammar' of the words in Wittgenstein's eccentric but handy sense of 'grammar'). You may shoot at any number of things, but what you shoot at is

always the target. In any case, the 'because' in 'I want it because it's good' indicates a genuine dependence: for the formula is not normally reversible. There is something pathological about 'It's true because I believe it' or 'It's good because I want it'. The latter phrase sounds less odd than the former, because ethical Protagoreanism is more fashionable (and more plausible) than its epistemic counterpart. But in either case if we take the phrase seriously it precludes deliberation. To decide on my wants (or beliefs) is not to ascertain them. Of course this is not a proof of the objectivity of the good and the true: for a philosopher such as Spinoza, for example, the objectivity of truth is grounded on something quite different, and the subjectivity of good is unaffected by its status as formal object of wanting: in either case, deliberation is merely illusory. But it does show that if we deny the dependence of wanting and believing on their formal objects, we must give up some deep and pervasive common beliefs about the existence and function of deliberation.

I said that connected with each presupposition there is a puzzle. The puzzle here is this. There is a temptation, going back to Socrates, to take the formal object of wanting for its real object. This results in the doctrine that no one ever (*really*) desires the bad. But there is no analogous temptation with regard to truth: in Plato, 'No one believes the false' is taken as a *reductio* of the views that imply it. And yet if the relation of truth to belief is the same as that of good to wanting, one would expect that the move would have the same plausibility in either case. What I really want is the good: if something that I actually pursue turns out not to be good, that simply means that I was wrong in thinking that this was the object of my pursuit. I have simply misidentified the good: but the good is what I wanted. Similarly if I believe what is false: what I *really* believed (the target of my mental state) was the *true*: if what I ostensibly believed turns out to be false, this is again a simple case of misidentification: I mistook that proposition for the truth. But what I really believed all along was *the truth*. Now why does that sound ridiculous?

One rather impressionistic hypothesis is this: wants are partly determined by beliefs without being identical with beliefs; thus it seems to make sense to identify a want while repudiating its 'component of belief'. In the case of belief, however, there is no reciprocation; and it would be obviously absurd to pretend to identify a belief while dissociating it from its belief components.

This needs amplification. What is the 'belief component' of a want? One fairly natural answer[1] is that it is the belief *that the object wanted is good*. Thus it may look as if we could explicate 'S wants that p' as (4):

(4) S wants that p and S believes that p is [part of the] good.

where the second conjunct of (4) can be dissociated from the first. If I am right, this is an illusion: that is the force of the claim that good is the formal object of wants. But if we try to establish a parallel 'analysis' of belief (5), it does not even *look* possible to dissociate the conjuncts:

(5) S believes that p and S believes that p is true.

This may be why there is a tendency to hypostatize the good though there is very little tendency to hypostatize the true. The unity of the true can take care of itself: logic is logic, and there is no need for hypostatization. But in so far as one has an initial prejudice in favour of the unity of the good, it seems easiest to justify it *via* the requirements of consistency on the *true*—precisely because the unity of the Good is on shakier ground. And to make the requirements of consistency applicable to the good, the simplest method is to make an object out of it: the requirements of consistency will then simply follow from the fact that propositions about the good will purport to express facts about an object and therefore be subject to the same constraints as other factual propositions. And now once you have hypostatized the Good, it is a small step to making it the real object of wanting. Any specific object of a want will then be interpreted as identified by the want with the Good.

I now pass to the second presupposition of my formulations: *possible objects of wants are possible objects of belief*. This is the standard view, which seems required by the fact that one sometimes gets what one wanted. If one is to be satisfied by the fact that what one wanted has come about, it seems that what one is now aware of having come about must be the very thing that one wanted to come

---

1   Belief may enter into desire in another way too. The desirability of any proposition p can be viewed as a weighted average of the desirabilities of all the complete 'novels' or world descriptions of which p is a part, where the weights are the probabilities attributed to each of these possible ways for p to turn out true. (See Richard C. Jeffrey, *The Logic of Decision*, New York: McGraw Hill, 1965, especially pp. 197–199.) Note that this scheme precludes the 'component of belief' in wanting from ever being separated out. See below (pp. 542–543) for some comments on this type of semantics for wanting.

about. Thus strictly speaking the objects of wants are propositions (however analysed), which are possible objects of belief. Propositions are also, by definition, the bearers of truth or falsity. In my scheme they are also therefore bearers of goodness or badness.[1] One speaks loosely of wanting objects, to be sure, but what one wants is *to have* them, and to have them almost always means something else, which varies according to the object; to eat it, or touch it, or to possess it—the shadow of enjoyment based on legal sanctions [e.g. marriage]. I leave aside the wants of creatures that are not generally supposed to have any propositional attitudes yet are supposed to have wants and perhaps 'know' things too. Though these are fascinating in themselves and in their relation to propositional attitudes, they are not relevant to my discussion. I am interested only in those states that *can* be said to have 'formal objects' specified by propositions, for only those are subject to assessments of rationality.[2]

This second presupposition about the objects of wants and beliefs might seem in itself to support the thesis I am attacking. The puzzle here is that there are relations among beliefs and their objects which do not seem to have analogues among wanting and its objects. Wants may be *satisfied* or not: what is the analogue of satisfaction for beliefs? The obvious candidate is *truth*: but we have already pre-empted *goodness* as the analogue of truth in its relation to wanting. The relation of truth to beliefs, therefore, seems to play a dual role in the analogy with wants: it is analogous to the relation of good to wanting, and also analogous to the relation of satisfaction borne by *truths* to wanting. Let us define a univocal sense of *satisfaction*, for a belief or a want, as *truth* of the sentence that specifies its object. And let us define *success* as the attainment, by each state, of its respective formal object. Then we get the following analogies:

(6) the *belief* that p is *successful* iff p is *true*.
    the *want* that p is *successful* iff p is *good*.
    the *belief* that p is *satisfied* iff p is *true*.
    the *want* that p is *satisfied* iff p is *true*.

1   This usage, I concede, is not a natural one. It commits one to some rather awkward turns of phrase. For example, it obliterates the difference between 'it is good that p' and 'it would be good if p', which are more natural if p is held to be true or false respectively. It might be smoother to speak of *states of affairs*, as being actual or non-actual, desirable or not desirable. I shall assume that nothing of importance hangs on these matters.

2   See 'How To Give a Piece of Your Mind (or the Logic of Belief and Assent)', *The Review of Metaphysics*, xxv, 1 (1971), 52–79.

A disanalogy is now evident: a single condition ensures both success and satisfaction of belief, but success and satisfaction for wants are independent.

This disanalogy lies at the root of one more urgent and more often noticed. If the belief that p is satisfied, no other truth can render it unsuccessful—in the sense just defined. But a want satisfied can be made unsuccessful by the conjunction of some unexpected other truth. I want two hundred pounds. My want is satisfied by my receiving compensation when my son is killed. This might be called the 'Monkey's Paw' phenomenon, after the short story by W. W. Jacobs in which a magical monkey's paw has the virtue of bringing about the realization of wishes in disastrous ways. It involves the relation of goods [and wants] to *truths*. In this it contrasts with the rejected parallels set out in (2) and (3) above (p. 538), which dealt only with criteria of internal consistency for sets of goods (or wants).

The Monkey's Paw phenomenon may tempt us to the view advocated by Jeffrey that the real objects of wants are not pro-positions after all, but total states of affairs (complete novels) of which the expression of the want specifies only a fragment. Hence, unless we specify a complete world (which is obviously impossible) our want statements are always elliptical, and can be defeated by the addition of conjuncts that negate some of the hidden condi-tions. Nothing of the sort is true for beliefs: their objects remain partial state descriptions, and it is trivial that no true partial state description may conflict with any other true partial state description.

The trouble with this suggestion is that it implies that although we can know what we believe we can never know what we want. For obviously no one can have a complete world description in mind when he expresses a want. So I might always discover more exactly what I wanted, at the same time as discovering that the actual world is not, after all, the world I wanted, even though it resembles that world in containing p, which is what I said I wanted.

Moreover this view fails to explain a more basic disanalogy: viz., that belief is conjunctible while wanting is apparently not. That is to say, if I believe p and believe q, rationality requires me to believe the conjunction (p & q). For I cannot be wrong in believing the conjunction if I am right in believing each conjunct, and con-versely if I am wrong in believing the conjunction I must have already gone wrong in respect of at least one of the conjuncts. Not

so for wants: I may desire *that I eat shrimp now* and desire *that I eat custard now* but have no particular desire *that I eat shrimp and custard now.* And intuitively this may be perfectly rational. On the semantics favoured by Jeffrey, it is an obvious move here to say that the real total object of either of my wants is my complete ideal novel: and if my complete ideal novel contains each conjunct, it must contain the conjunction. This is the 'ideal world' conception of rationality which I have rejected: and although it is not entailed by the 'total world' semantics of wanting (for 'the' ideal world might not be unique) it fits in naturally with it. This together with its intrinsic implausibility inclines me to reject it.

The most sensible course in the face of these problems about satisfaction is to insist that whenever a person wants p and p comes about, his want is thereby satisfied. From which it does not follow, of course, that the *person* is now happy: that *he* is 'satisfied'. He might have other wants, indeed wants induced by the very satisfaction of the first, and be more miserable than before. But that is not because his initial wants have not been satisfied: it is in spite of the fact that they have. We have seen how the monkey's paw phenomenon can be understood in terms of the dissociation of success and satisfaction, where the *person's* satisfaction depends not only on the satisfaction (in the defined sense) of his desires but on their continuing to be seen as successful. The point about conjunctibility is equally well explained by the facts exhibited in schema (6). A conjunctive want must be satisfied if and only if its conjuncts are. But contrary to the case of belief this does not entail that it must be *successful* if its conjuncts are. That is precisely the point at issue: whether a conjunction of goods must itself be good. It is not a point that should be legislated a priori by semantics.

The point applies to second-order beliefs and wants. In a paper on 'Happiness', A. Kenny has quoted Russell as maintaining that 'to be without some of the things you want is an indispensable part of happiness'. Kenny labels this 'ω-inconsistency', a term applied by logicians to infinite expressions of the form:

(7) $(\exists x) \sim Fx \ \& \ Fa_1 \ \& \ Fa_2 \ldots \& \ Fa_n \ldots$

Kenny remarks that 'modesty seems to demand that we should hold ω-inconsistent beliefs, e.g. that we should believe that some of our beliefs are false.'[1] Actually the peculiarities of the cases

1 A. Kenny, 'Happiness', in J. Feinberg, ed., *Moral Concepts* (Oxford, 1970), p. 45.

Kenny is concerned with have nothing to do with expressions of the form of (7). Allowing 'O' to stand either for belief or for want, they have the form:

(8) Op_1 & $Op_2 \ldots$ & Op_n & $O(\exists p)$ $(Op$ & $\sim p)$.

They might be better termed 'Moore-inconsistency', since (8) is the form of 'Moore's paradox'. Whether and why Moore-inconsistency in beliefs is genuine inconsistency has been a matter of debate. In the terms I have defined, I propose the following criterion:

(9) A person is inconsistent in respect of a set of states O if their specification alone guarantees some *unsuccess* among the members of O (cf. p. 537 above).

By this criterion, what 'modesty demands' is inconsistent. But Russell's conception of happiness does *not* commit him to inconsistency. If we interpret O in (8) above as referring to Russell's wants, we can see that (8) guarantees only that his wants are not all *satisfied*. (It does this by guaranteeing that his second-order want *is* satisfied provided its scope is not restricted.) But this has no bearing on whether Russell's wants are *successful*. This is evident from the definition of 'success' for wants in (6) above. Consequently 'Moore-inconsistency' is not a sufficient condition of inconsistency for wants. It is for beliefs.

Thus far I have argued that a demand for consistency of wants modeled on consistency of beliefs is gratuitous, and that viewing possible objects of wanting as possible objects of belief leaves the question open. I also noted at least two types of case in which intuitively a disanalogy exists between rational wants and rational beliefs. I now want to resume the discussion of utilitarianism, and give some more positive considerations in favour or allowing conflicts of goods and (hence) rational conflicts of wants.

For the sort of classical Utilitarian theory I have described, indifference, like preference, in terms of which it is defined, is a transitive relation: given any two alternatives A and B, if A and B are indifferent, then any third alternative indifferent with A is indifferent with B. Together with the postulate of *connexity* (p. 536 above), this gives excessively strong conditions on rationality.

To see this, consider the case of the Fairly Virtuous Wife. I tempt her to come away with me and spend an adulterous weekend in Cayucos, California. Imagine for simplicity of argument

that my charm leaves her cold. The only inducement that makes her hesitate is money. I offer $1,000 and she hesitates. Indeed she is so thoroughly hesitant that the classical decision theorist must conclude that she is *indifferent* between keeping her virtue for nothing and losing it in Cayucos for $1,000. (Actually this turns out to be quite a difficult thing to *establish*: for technically a subject is said to be indifferent between two alternatives if he chooses each apparently at random about half of the time. But let us suppose that this practical problem is solved.) The obvious thing for me to do now is to get her to the point of clear preference. That should be easy: everyone prefers $1,500 to $1,000, and since she is indifferent between virtue and $1,000, she must prefer $1,500 to virtue by exactly the same margin as she prefers $1,500 to $1,000: or so the axioms of preference dictate. Yet she does not. As it turns out she is again 'indifferent' between the two alternatives. The classical Utilitarian is forced to say that she is incoherent, because she violates his axioms of rationality. Her preferences are represented by the following, which obviously violates the transitivity of indifference:

$$\$1,000 = V = \$1,500, \text{ but } \$1,500 > \$1,000.$$

I would prefer to say that the alternatives considered are *incomparable*. The proper representation of her preferences is this:

$$\sim(\$1,000 \geqslant V) \ \& \ \sim(V \geqslant \$1,000) \ \& \ \sim(\$1,500 \geqslant V) \ \&$$
$$\sim(V \geqslant \$1,500) \ \& \ \$1,500 \geqslant \$1,000 \ \& \ \sim(\$1,000 \geqslant \$1,500).$$

We have dropped connexity, but she is not irrational. Other examples are easy to find. Any of the alternatives which Mill describes as 'qualitatively different' will do.

So will that traditional bugbear of Utilitariansim: the clash between Justice and Utility. If a sacrifice of utilities is required, then the smallest possible sacrifice is evidently better than the larger. Yet suppose the sacrifice of two innocent individuals maximizes Utility but clashes with the demands of Justice. If no rational decision appears possible, the case is not made evidently simpler by slightly reducing the sacrifice or slightly increasing the Utilities. So the value of Justice is not just another source of utiles, evenly balanced against the sum of other utiles.

In this area, there are no knock-down arguments. It is easy for the advocate of the classical view to explain away all the counter-examples. The case of Justice has been discussed so much that

18

there is no point in trying to rehearse the arguments here; in the more modest case of the Fairly Virtuous Wife, we can say she changes her mind very rapidly; or that the apparent value of money in utiles undergoes a rapid inflation under the pressure of guilt. Or the alleged incomparability may simply be a consequence of the high degree of uncertainty in the situation (How pleasant is Cayucos? Will my husband find out? And so forth.) In epistemic matters too there may be cases where the uncertainty makes the canons of rationality inapplicable: witness the paradoxes generated by the Principle of Indifference in Probability Theory. There are matters on which no rational guess is possible simply because there is no antecedently given number of possible (hence *a priori* equiprobable) alternatives. Sure enough. But all these explanations are *ad hoc*. Why insist? The reason for insisting seems to be two-fold. First, it appears to be demanded by the behaviourism that belongs to the same philosophical tradition as Utilitarianism. Second, there seem to be no alternatives for the explication of rationality; thus one might as well be content with a tidy theory and let the anomalous cases fall where they may.

Both these considerations, however, are mistaken. The kind of behaviourism that Utilitarians and Decision Theorists need is no strong form of operationalism. Many 'wants', in the weak sense, need not be reflected in any action, as I noted before. They are simply listed on the agent's desirability scale and may or may not result in action if the right combination of beliefs and other circumstances presents itself. From the point of view of the theory, wants do not cease to exist when they are not dominant. It is only wants-all-things-considered that are supposed to be tied to actual choices. The theory concedes the existence of ordinary wants, and those may conflict without the conflict being manifested in an impossible pattern of behaviour (both choosing A and not choosing A).[1] Let us then look for an account of incomparable 'everything-considered' wants which both gives them empirical content within a theory as behavioristic as the one I have been discussing, and avoids the demand that such pairs of wants result in impossible behaviour.

1   Expressed in terms of 'good', this is the view that relative goods (or *'prima facie* goods') can conflict: but these are not real conflicts of the 'absolute good' with itself. *Prima facie* goods have been illuminatingly compared to statistical probabilities: a proposition may be highly probable in relation to evidence *e*, but improbable relative to another body of evidence *e'*. This does not mean, needless to say, that there are conflicts within the Truth. (See D. Davidson, 'How is Weakness of the Will Possible?', in J. Feinberg, ed., *Moral Concepts*.)

If we can find such an account there should be no plausibility left to the classical theory at all, nor to the Unity of the Good.

Such an account may I think be found around a notion of *multidimensional value*. A dimension, for my purposes here, is a scale along which comparative measurements may be made but which is incomparable with any other such scale of dimension. An increase in one dimension will leave the other dimensions unaffected: so in the case of the Fairly Virtuous Wife the increment of money, while clearly perceptible in its own dimension, makes no difference in the balance against another dimension of value. In each dimension there will be *prima facie* and *everything-considered* wants. Either type of want, however, will belong exclusively to a single dimension only if it is *basic*: that is, if it is not had for the sake of some other want on the basis of some practical argument.[1] Most wants that are immediately related to action are derived ones, and it is about these that conflicts arise. Thus while there can be no two inconsistent everything-considered wants within a single dimension, two such wants can conflict if they belong to different dimensions, and they can be directly contradictory if they are derived by argument from wants in different dimensions. (Note that this definition does not exactly fit our common-sense notion of spatial dimension. Extension along one co-ordinate is by definition independent of extension along another, but the notion of spatial *interval* does allow comparisons between lengths in various dimensions. This is because we use the handy convention of *rigidity under transport*. Nothing corresponds to that in dimensions of value. There no is 'transport'.)

Empirical evidence for the adequacy of such an account would be found in the sort of pattern exemplified in the FVW case, together with the possibility of producing sensible increments within any single dimension. One can imagine experiments somewhat on the lines of those conducted some years ago by Davidson, Suppes and Siegel at Stanford. From those experiments it appeared that subjects did indeed exhibit, in simple experimental circumstances, the patterns of choice predicted by subjectivist Decision Theory.[2] One might set up clashes of Utility and Justice, for example, to see whether clear differences between the justice

1   For a discussion of basic intrinsic vs. derived wants, cf. Gilbert Harman, 'Toward a Theory of Intrinsic Value', in the *Journal of Philosophy*, lxiv, 23 (1967), 792–804.
2   D. Davidson, P. Suppes and S. Siegel, *Decision Making, an Experimental Approach*; Stanford, 1957.

values alone, or utility values alone, of different arrangements, leave the dilemma unaffected.

Here let me take up a remark I made earlier: that on the unidimensional view there is no place for regret. On the multi-dimensional view of value, it is clear how the presence of incompatible wants might naturally lead to regret. If I am forced to choose between two deals, then when I choose the most profitable it is irrational to regret the other deal: for this one gave me whatever the other would have given me, and then some. But if I am forced to choose between a Higher and a Lower pleasure, then whichever I choose I can regret the loss of the other, for what I got from the one I chose was qualitatively different—was a value in a different dimension—from what I have lost in not having the other. Here there is absolute loss. Since there is no universal *hedon*, of which I can say I would have got so many from the alternative rejected and got so many (perhaps more) from the one chosen, the concept of simply maximizing utilities no longer makes sense. Goods conflict absolutely.

It might be objected that finding a place for the notion of regret is no argument in favour of a view of rationality, for regret is never rational. 'He who repents is twice unhappy or twice impotent' (Spinoza, *Ethics* iv. 54). Actually I am inclined to agree; but that is not to the point. The question of what constitutes a 'rational' attitude to loss or for that matter to any inevitable unpleasantness is a problem not to be solved by quasilogical considerations. The point is that a multi-dimensional value scheme *recognizes* the possibility of inevitable loss. On a single-dimensional scheme the only loss there can be is a failure to maximize utility. Doubts about the rationality of regret indicate something which it has been partly my aim to establish: that as soon as we get beyond some simple rules for consistency of belief and for suiting the means of action to its ends, 'rationality' becomes something on which we have disparate intuitions governed by a variety of not particularly well-connected prejudices.

Let me recapitulate. I began by arguing that we should not prejudge the nature of rational wanting by reference to the criteria for rational belief, since the reasons for thinking that true propositions cannot be logically inconsistent (in contrast for example to *possible* propositions) do not carry over to propositions about what is good. I then considered some disanalogies between wanting and believing which arise within the point of view that considers them

as closely parallel as possible in their relations to their formal objects. Finally I gave reasons for preferring a version of value theory which is multidimensional and allows for absolute conflicts of goods and wants. But I have said practically nothing about what *are* the principles of rational wanting.

Is *any* set of conflicting wants acceptable whatever the source of the conflict? In particular, is the conflict allowed to be logical or merely practical? If it is practical only, there is after all no dis-analogy with the case of belief: for we may without irrationality believe propositions that happen to be practically incompatible. On the other hand, if logical conflicts are allowed (if it can be rational to want that p and want that q where q is logically in-consistent with p), then the point about loss and regret is without force. Even on a single scale of utility, if I am allowed to want contrary states of affairs, I can also want the state of affairs of my having *both* of two goods known to be inconsistent: thus I can indeed regret the loss of the one I have not got.

This last suggestion, however, is absurd. I do consider myself committed to a sensible modicum of behaviourism: a want, even if it is not an 'all-things-considered-want', must be such that under some circumstances it *might* issue in action. But nothing could count as an attempt to bring about a state of affairs known by the agent to be self-contradictory. There might be rapid alterations of behaviour aimed at causing and preventing a given state of affairs, but this is not the expression of a single want. Within each dimen-sion and in so far as we are dealing with *basic* wants not with *derived* ones, we should accept the principle of rationality that pre-cludes wanting both p and ∼p. My reason here is this. Decision theorists have boldly assimilated criteria of ascription and criteria of rationality: on the classical model, any instantaneous preference ranking that might be exhibited in an actual series of choices—criterion of ascription—is thereby deemed coherent—criterion of rationality. At the most basic level this is quite correct, simply because a minimal concept of rationality is embedded in our very concepts of action, belief, and want; our criteria of attribution must reflect this.[1] The concept of multidimensional value allows wants to be rationally incompatible, because the criteria of ascrip-tion of wants belonging to one dimension are not relevant to the criteria of ascription belonging to another. Thus one can want

[1]   See 'How to Give a Piece of your Mind . . .' for some discussion of the relation between criteria of ascription and criteria of rationality.

states of affairs which in conjunction with other factual beliefs
turn out to be contraries. And this is sufficient to differentiate the
rationality of wants from that of beliefs in the way that I set out to
do. In the case of beliefs it is not rational to believe p and believe q,
if in virtue of some other [set of] beliefs p and q are incompatible.
That is, it is irrational to believe an inconsistent conjunction. But
it is not irrational to want two states of affairs which are incon-
sistent in the light of a third that one believes to hold. A corollary
is the disanalogy already noted: that rational belief is 'conjunctible',
whereas rational wanting is, in general, not.

I leave many questions unanswered. In particular there are a
number of mixed putative principles about wanting and believing
which it would be interesting to explore. For example the question
of whether wanting p and believing that p entails q commit one to
wanting q; or whether—as argued by Wallace in a recent paper[1]
the rational man wants everything he believes. Rather than pursue
these problems or list any more, I shall conclude with a couple of
remarks about the objectivity of value.

A consequence of my view is the resuscitation of the 'Fact-
Value distinction'. If I am right, facts and values have different
logical structures. And there are two problems which this may seem
to raise.

First, since a simple trick of language can change any expression
of want into an expression of belief (about the good), how can
wants and beliefs differ in general in their logical behaviour?
Actually this is no problem. 'Good' is an operator which although
it applies to the objects of belief is not truth-functional: it is not to
be expected that 'it is good that . . .' should function as does 'it is
true that . . .' any more than 'it is possible that . . .' or 'it is prob-
able that . . .'.

The second problem is a little more tricky. I recognize the tempt-
ation to say[2] that only that can be *objective* to which the criteria of
consistency apply. Why? Because what is objective must be in the
world, and what is in the world cannot be contradictory. There-
fore, one might conclude, a view which allows that the good is not
bound by the criteria of consistency must deny the objectivity of
value.

The issue of objectivity is difficult and unclear; but I think it is at

1   John Wallace, 'Propositional Attitudes and Identity', *Journal of Philosophy*,
     lxvi, 6 (1969), 145–152.
2   Cf. B. A. O. Williams, 'Consistency and Realism', Aristotelian Society
     Supplementary Volume 40 (1966).

least clear that the objectivity of 'good' is not to be confused with its hypostatization, which as we saw is common, and leads directly to the unacceptable dogma. I think the inference just sketched is based on a confusion, dispelled by much the same consideration as that of the previous paragraph. The whole point of my thesis has been that propositions can be simultaneously *good* without being simultaneously *true*: thus one can maintain that values *are* objective, simply by pointing out that the *facts* about value which are objective (and therefore need, as propositions, to be all true together) are the facts that various propositions *are good*: for that, the propositions in question themselves do not need to be true. Hence they do not need to be consistent either. This view does not, then, commit one to subjectivism in matters of value. The role of subjectivity is in determining preferences among goods which belong to different dimensions, that are *objectively incomparable*. Only subjective preference can guide us in a choice such as Yeats', between 'Perfection of the Life, or of the Work'. The objective facts are that both are indeed perfections, that they conflict absolutely—that is, can never be realized in one man—and that 'the intellect of man is forced to choose'.

UNIVERSITY OF TORONTO