CHAPTER 30

S. DANIELSSON

HARMAN'S EQUATION AND NON-BASIC INTRINSIC VALUE

1.

Gilbert Harman (1967) argued that the standard notion of intrinsic value is characterised by the equation

$$(1) \qquad I(A) = \sum_{A \Rightarrow B} v(B)$$

where $I(A)$ is the intrinsic value of A, $v(B)$ is the basic intrinsic value of B and $\Rightarrow$ is the relation of entailment. The intrinsic value of an alternative equals the sum of the basic intrinsic values of the alternatives entailed by it.

Thus according to Harman, intrinsic value is a) additive and b) in a sense, derived from basic intrinsic value. I have argued elsewhere that the latter of these two assumptions may shed some light on some standard problems with the former.[1] Here I shall assume that basic intrinsic value is defined, and that the question is how far Harman's equation can be used to calculate non-basic intrinsic value.

This question was raised by Warren Quinn (1974), who modified Harman's idea in order to make it possible to speak of the intrinsic value also of A's which entail no basically evaluated B, but entail, for instance, disjunctions of such B's. Edward Oldfield (1977) argued that Quinn's modification did not work and suggested another. Erik Carlson (1997) suggested a third one. Here I roughly agree with Oldfield against Quinn, and with Carlson against Quinn and Oldfield, but in certain respects I disagree with the three of them.[2]

---

[1] Danielsson (1997)

[2] The first version of this paper was written in 1978. Erik Carlson (1997) made me reconsider some of its arguments. I am grateful to him, as well as to Jan Odelstad and Krister Bykvist, who pointed out several difficulties.

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 371-378.

2.

Harman says that what may have value is that such and such is the case. He explicitly rejects discussing what a phrase such as 'that the cat is on the mat' may refer to.[3] Quinn and Oldfield assume that the bearers of intrinsic value are propositions. They identify propositions with sets of possible worlds. I shall essentially follow them in this respect, though, like Carlson, I would like to say that A, B, C, etc. are (possible) states of affairs rather than propositions. In any case, they are assumed to be identified by that-phrases. I shall also assume that there is a basic set U (the tautologous state of affairs), and that the variables 'A', 'B', 'C', etc. refer to subsets of U. Hence $\Rightarrow$, the relation of entailment, will be identified with $\subseteq$, set inclusion, and we shall be able to speak of the negation $\neg A$ of a state A, which is the complement in U to A, the disjunction $A\cup B$, which is the union of A and B, and the conjunction $A\cap B$, which is the intersection. That states of affairs are thus identified with sets of possible worlds is not important for the argument; it is simply a convenient way to model entailment and the notions of disjunction, conjunction, etc.

More important than these technical details is the question of how the distinction between intrinsic value and *basic* intrinsic value should be understood.

Harman argued that the standard account of intrinsic value forces its adherents to assume that there is basic intrinsic value. For the standard account, according to Harman, says that the intrinsic value of A is the sum of the intrinsic values of what is entailed by A. But that does not work if, for instance, both A and *A and B* have intrinsic value, for then the standard account entails that the intrinsic value of *A and B* is the intrinsic value of *A and B* plus the intrinsic value of A.

The force of this argument be what it may, the notion of basic intrinsic value obviously corresponds to the notion of *intrinsic value as a whole* in *Principia Ethica*,[4] where this is contrasted with *intrinsic value on the whole*, which is what is usually meant by 'intrinsic value'. It is not all that easy, however, to provide a satisfactory definition. One might try with *underived* value, saying that the basic value of A is not dependent on the value of something else. This seems satisfactory in many cases. Thus, for instance, the hedonist would say that the value of the compound alternative *that there are happy egrets and there are stones* is not basic, since its value is derived from the value of the two conjunction members *that there are happy egrets* and *that there are stones.* And we could all agree that extrinsic value is not basic since it is, by definition, value as a means to an end, and, as such, dependent on the value of the end.

[3] We should perhaps distinguish between 'It is good that the cat is on the mat' and 'It would be good if the cat were on the mat', the former entailing, implying, presupposing or at least suggesting, that the cat is actually on the mat, the latter that it is not. The ascription of value to the cat's perhaps just possibly being on the mat, however, seems to be the same in both cases, and this is what we are interested in here, although there seems to be no natural expression for it in ordinary language.

[4] Cf. G. E. Moore *Principia Ethica* (Cambridge: Cambridge University Press, 1903), section 129.

Some would like to say, however, that the negative value of *Alec is enjoying the thought that somebody is in great pain* is basic, although it is dependent on the negative value of the alternative *that somebody is in great pain.* Perhaps we could say that the characteristic feature of basic value is that it is not *entailed* by the value of something else. For it seems natural to say that while the hedonistic ascription of value to the conjunction above is entailed by the ascription of value to the conjunction members, the ascription of negative value to Alec's enjoying the suffering of somebody else is not entailed by the ascription of negative value to the suffering. In the latter case, the dependence is moral, not conceptual.

As the hedonistic example above illustrates, not every intrinsic evaluation has to be basic. It is less clear whether any basic evaluation can be non-intrinsic. To say that something is intrinsically good is to say that it has a positive value which is independent of the actual circumstances and consequences. To say that something is good in a non-intrinsic sense, then, would be to argue that it is good because of some presumed fact about the circumstances or consequences of the evaluated alternative. It is not obvious that you cannot base your value-argumentations ultimately on premises of that form.

But I am not concerned here with the question of whether every evaluative system must contain some basic intrinsic evaluation. I am only trying to clarify the distinction between basic and non-basic intrinsic value. The question is complicated by the fact that, as far as the Harman equation is concerned, the intrinsic value of a basic alternative need not be the same as the basic value of the alternative. In Quinn's, Oldfield's and Carlson's theories, this possibility is blocked; therefore, they can, and indeed do, drop the function v from Harman's approach. They need only speak of a set of *basic alternatives*, which are those basically assigned intrinsic value, and this is, in fact, the approach they take.

These authors can reduce the distinction between basic and non-basic intrinsic value to the difference between basic and non-basic alternatives because they assume that no basic alternative entails any other. If, for instance, *A* and *A and B* are both basic alternatives, and we have just one kind of intrinsic value, we are back to the situation which inspired Harman's original argument. Unfortunately, there are easily described evaluative systems with this particular feature. A retributivist with hedonistic inclinations, for instance, may argue that the whole composed by a criminal's committing a crime and his suffering punishment for the crime has a positive basic value, while the parts, the committing of the crime and the suffering of punishment, both have a negative basic value which is so great that the total intrinsic value of the whole is negative. And an egalitarian hedonist may argue that what has basic intrinsic value (positive or negative) is either that a certain individual is happy (or unhappy) to a certain degree, or that there is a certain difference between the degree of happiness of one individual and that of another. (In fact, it seems to me that no axiology with any plausibility at all can assume that no alternative with basic intrinsic value entails any other such alternative.)

Thus if we want a more general theory of intrinsic value we seem to need the notion of basic intrinsic value as a basic notion.

3.

Let us assume then that there is a set U, the subsets A, B, C,... of which are the states whose intrinsic value is at stake.

We assume that **V** is the set of basically evaluated alternatives, the alternatives for which the real-valued function v is defined.[5] To ensure that the sum of basic value mentioned in the Harman equation is well-defined, we assume that every non-empty (consistent) conjunction of members of **V** is finite.[6]

The Harman equation (1) now assigns positive or negative intrinsic value to an alternative if and only if it entails some member of **V**. On the one hand, this seems to be too restrictive. There may be non-basic alternatives which we would like to say are intrinsically better than certain others, and some alternatives which we would like to say are intrinsically good or intrinsically bad although they entail no member of the set **V**. It seems plausible to assume, for instance, that a disjunction $A \cup B$, where A and B both have a positive basic value, has itself a positive value, even if $A \cup B$ entails no member of **V**.

On the other hand, an alternative may also entail a member of **V**, and thus get a definite intrinsic value by the equation, although this value obviously differs from what seems to be intuitively plausible. Using hedonism as an illustration, assuming that pleasure and pain can, in some way or other, be measured, and that basic intrinsic value is assigned only to the precise degree of pleasure or pain of an individual, we may consider the alternative *that Alec is enjoying himself to degree 2 and Bertie is suffering to degree -1 or suffering to degree -10*. The Harman equation would give that alternative the intrinsic value of *Alec enjoying himself to degree 2*, since that is the basic value entailed. But clearly the compound alternative should be worse.

Hence the Harman equation in its original form must have a very restricted field of application. It may be applicable to members of **V**, conjunctions of such members, and alternatives formed by conjoining to such conjunctions alternatives which neither entail, nor are entailed by, members or conjunctions of members of **V**. In particular, the equation seems to be applicable to the total intrinsic value of a possible world, which hence would be the sum of the basic intrinsic values of alternatives that are actually realised in the world in question. This fits well with the ingenious dictum in Chisholm and Sosa (1964) that 'what is intrinsically good is what rates the universe a plus'. But Harman's main idea could be made more generally applicable, and this is what Quinn, Oldfield and Carlson have tried to do. Here is an alternative way of doing it.

---

[5] We usually consider the evaluations from an external perspective. We might as well take an internal one and say instead that **V** is the set of basically *valuable* alternatives.

[6] Notice that this does not mean that **V** is finite. There might well be an infinite number of levels of intrinsic value. But it means that there is a finite number of relevant aspects of the value of an alternative, in the sense that the intrinsic value of an alternative can always be accounted for by reference to a finite number of entailments.

4.

Compared to several earlier theories of intrinsic value, Quinn's theory seems to take a considerable step forward by assuming that an alternative may have an *indefinite* intrinsic value. Much logic of preference would have been happily undone if we had generally recognised this possibility. Such alternatives, however, are assumed by Quinn and Oldfield to have a definite *intrinsic value at a world* that shifts from one possible world to another. As Carlson emphasises, this seems to contradict a standard assumption about intrinsic value; the version of Harman's theory sketched below will *not* assume that intrinsic value varies from one possible world to another. The idea that intrinsic value may be indefinite, however, will be retained, and, like Carlson, we shall let the value of such an alternative be represented by a *set* of numbers instead of by a number.[7] The basic idea is simply that if A and B are both basic, then the value of $A \cup B$ is (represented by) the set $\{v(A), v(B)\}$.

Since we have assumed that basic alternatives may entail each other, this idea has to be somewhat restricted, however, lest it entail that the value of a basic alternative A is represented by a set containing every $v(B)$, where B is a basic alternative entailing A. In order to avoid this, we define an extended basic value-function v^*, which assigns a set of numbers to every member of the union closure of $\mathbf{V}$, the set of all finite unions of members of $\mathbf{V}$, which we call $\mathbf{V}^*$:

(2) $v^*(A) = \{v(B)$*: B is in* $\mathbf{V}$, $B \subseteq A$ *and there is no C in* $\mathbf{V}$ *such that* $B \subset C \subseteq A\}$.

Thus if the function v assigns a definite basic intrinsic value to the members of $\mathbf{V}$, then v^* assigns an indefinite basic intrinsic value to disjunctions of these basic alternatives (including the one-term disjunctions of the basic alternatives themselves). The basic value brought into the world by *A or B* being the case is the basic value of A or the basic value of B. Notice that the restriction in (2) makes this basic idea applicable only to true disjunctions, so to speak, for if B is included in A, and $A \cup B$ is hence actually identical to A, then $v^*(A \cup B)$ will contain just $v(A)$, not $v(B)$. That which already has a definite basic value does not need an indefinite one. Or, to put the matter differently, the value of an alternative turns into a more indefinite one only if the alternative itself turns into a more indefinite one.

We then say that *B is a maximally v-definite consequence of A* if and only if $A \subseteq B$ and either B is in $\mathbf{V}$ or (*i*) B is in $\mathbf{V}^*$ and (*ii*) there is no C in $\mathbf{V}^*$ such that $A \subseteq C \subset B$.

It is evident that when we have listed the maximally v-definite consequences of A, then we have listed, as precisely as possible, which basically valuable things will

[7] Carlson actually uses *intervals* of real numbers. Since intervals can be taken to be sets of numbers, this may be considered a special case of the set-approach.

necessarily be in the world if A is true. And we take their added value to be the intrinsic value of A. Thus our new version of Harman's equation will be:

(3) $I(A) = \sum v^*(B)$

B is a maximally v-definite consequence of A.

We extend the relation > to sets of numbers, assuming that $v^*(A) > v^*(B)$ *if and only if* $x>y$ *for every* x *in* $v^*(A)$ *and* y *in* $v^*(B)$. We make the analogous assumption about the function I, and we make it possible to add indefinite values by assuming that $v^*(A) + v^*(B)$ is the set of all sums $x+y$ where x is in $v^*(A)$ and y in $v^*(B)$.

We assume further that A is intrinsically better than B if and only if $I(A) > I(B)$, and that A is intrinsically good (bad) if and only if $I(A)>0$ ($I(A)<0$). Given this, it seems natural also to assume:

(4) *If* $v(A)>0$, *then not:* $v(\neg A)>0$; *if* $v(A)<0$, *then not:* $v(\neg A)<0$.

(5) $v(U) = 0$.

If it is good (bad) that A, then it is not also good (bad) that *not A*. To be good (bad) is to be better (worse) than the tautology.[8]

It is evident that this version satisfies the demands on the treatment of disjunctions which were formulated at the end of section 3. Let us assume that the basic intrinsic value of *Alec is enjoying himself to degree 2* is 2, the basic intrinsic value of *Bertie is suffering to degree -1* is -1, and the basic intrinsic value of *Bertie is suffering to degree -10* is -10. The maximally v-definite consequences of *Alec is enjoying himself to degree* 2 *and Bertie is either suffering to degree -1 or suffering to degree -10* are (i) the tautology, (ii) *Alec is enjoying himself to degree 2*, and (iii) *Bertie is suffering to degree -1 or Bertie is suffering to degree -10*; the v*-values of these are {0}, {2} and {-1, -10}, and hence the intrinsic value of the alternative at stake is {1, -8}.

5.

This version is simpler than those of Quinn, Oldfield and Carlson. It is also more general, since it does not assume that the basic alternatives are logically independent of each other. Apart from these general features, the most striking particular deviations from Carlson's theory are probably that in the present one *every* member of $I(A)$ has to be greater than *every* member of $I(B)$ for A to be better than B, and that the value of *A and B* has no particular role in the calculation of the value of *A or B*.

---

[8] It certainly may be questioned whether the value of the tautology is really comparable to the value of anything else. But the main point of these assumptions is to make the numerical machinery work more efficiently.

Regarding the former, Carlson presents a couple of examples intended to show that the version chosen here is not a reasonable one. In the most important case, we are invited to compare *A or B* to *B* when *A* is better than *B* (and *A* and *B* are both basic). Is it not obvious that, since *A or B* involves the possibility of *A*, which is better than *B*, that the disjunctive alternative is also a better one? I do not think so. Perhaps it could be argued that, given some kind of choice situation in which we in some sense have to choose between *A or B* and *B* (it is not very clear how such a situation should be described), then choosing *A or B* is a preferable choice. But it is not obvious that 'intrinsically better than' should be equated with the formulation in terms of choice.

And if we assume that *A or B* is better than *B* in case *A* is better than *B* (and *A* and *B* both are basic), then our theory of intrinsic value is hardly compatible with hedonism.[9] For hedonism, strictly speaking (or perhaps we should say one strict kind of hedonism), is the assumption that *A* is intrinsically better than *B* if and *only* if *A* contains, that is, entails, a greater balance of pleasure over pain than *B* does. And it is obvious that *Alec is enjoying himself or Bertie is suffering* does *not* entail a greater balance of pleasure over pain than does *Bertie is suffering.* Perhaps we could say that the former alternative *might* bring positive value into the world by being true (*might* render the Universe a plus), while the latter certainly will not. But hedonism as it is usually conceived assigns positive value to pleasure, not to possible pleasure. As far as *Alec is enjoying himself* and *Bertie is suffering* are concerned, there is more pleasure in the world if the former is true than if the latter is true. But it is not the case that, as far as *Alec is enjoying himself or Bertie is suffering* and *Bertie is suffering* are concerned, there is more pleasure in the world if the former is true than if the latter is.[10]

It also seems to me that if, when *A* is better than *B*, we do not say that *A or B* is *better than B*, but say instead simply that *A or B* is *at least as good as B*, then we get a terminology which makes it possible to make some finer, though still very simple, discriminations. I think we should consider the possibility that *A* is intrinsically at least as good as *B* if and only if $x \geq y$ for every x in $I(A)$ and y in $I(B)$. That means, of course, given the assumption of 'better than' above, that we cannot adopt the equivalence between *A is better than B* and *A is at least as good as B and B is not at least as good as A*, which so often appears in the literature. But the equivalence seems to be a necessary one only if the betterness-ordering is a weak order. Nor will *A is at least as good as B* be equivalent to *A is better than B or equal in value to B.*

The notion of equality in intrinsic value actually appears to be a much more problematic one than the notion of betterness. There seem to be a number of more or

[9] Some people may be inclined to say that if our theory of intrinsic value is incompatible with hedonism, then so much the worse for hedonism. But although I am sure that hedonism is wrong, I do not think that it is wrong for logical reasons.

[10] We certainly can imagine a sort of hedonism which assigns basic intrinsic value also to the disjunction. Then the disjunction might be intrinsically better than the second disjunction member. So this kind of hedonism is compatible with our theory of intrinsic value, while what here is taken to be hedonism proper is not compatible with the alternative theories.

less plausible candidates ranging from *neither better nor worse than* and upwards in logical strength. One of the best, perhaps, says that *A* has the same intrinsic value as *B* when something is intrinsically better than *A* if and only if it is intrinsically better than *B*, and intrinsically worse than *A* if and only if it is intrinsically worse than *B*.[11] And to incorporate this idea into our theory smoothly, we can say that *A is equal in intrinsic value to B if and only if the maximal member of I*(*A*) *equals the maximal member of I*(*B*) *and the minimal member of I*(*A*) *equals the minimal member of I*(*B*). (The present theory will hence, on this point, agree with Carlson's.)

As to the role of the value of *A and B* in the evaluation of *A or B*, Carlson argues, like Oldfield, that in some cases *A or B* is better than *C* because *A and B* is better than *C*. For some reason, they assume that among all subalternatives of *A or B* precisely the subalternative *A and B* (if that is consistent) should have a special standing. Carlson argues that *A and B* is a 'realisation' of *A or B*, implying that the value of an alternative is dependent on the value of its 'realisations'. But it is in no way clear why *A and B* is a 'realisation' of *A or B* in a sense different than for any subset of *A or B* whatsoever.

REFERENCES

Carlson, Erik: 'The Intrinsic Value of Non-Basic States of Affairs'. *Philosophical Studies* 85 (1997): 95-107 [* pp. 361-70 of this volume].

Danielsson, Sven: 'Harman's Equation and the Additivity of Intrinsic Value'. In Lindahl, Needham, Sliwinski: *For Good Measure: Philosophical Essays Dedicated to Jan Odelstad on the Occasion of his Fiftieth Birthday* (*Uppsala Philosophical Studies*, 46 (1997)): 23-34.

Halldén, Sören: *On the Logic of Better*, Library of Theoria 1957.

Harman, Gilbert: 'Toward a Theory of Intrinsic Value'. *Journal of Philosophy* 64 (1967): 792-804 [* pp. 349-60 of this volume].

Oldfield, Edward: 'An Approach to a Theory of Intrinsic Value'. *Philosophical Studies* 32 (1977): 233-49.

Quinn, Warren: 'Theories of Intrinsic Value'. *American Philosophical Quarterly* 11 (1974): 123-32.

[11] This idea, though not for *intrinsic* betterness, is used already in Halldén (1957).