

# An Empirical Theory of Value

James E. Petts

## Table of Contents

Acknowledgements	2
Introduction	3
I. Why an empirical theory?	4
II. Value and life	6
III. Decisionmaking, orders of value, and pleasure	10
IV. The family of evaluative concepts	15
V. Value vagueness, and the incoherency of incommensurability	18
VI. Acts, rules and decisionmaking method	23
VII. Why the good of all is the right for each	25
VIII. Justice, law and punishment: dealing with the irrational	28
Conclusion	31

### **A note on version numbers:**

*Academics have much to learn from computer programmers in terms of revision numbering. Computer programmes are customarily given a non-integer version number to enable revisions and modifications to be tracked effectively. Thus, when errors are spotted in the programme, everybody can be clear as to the versions in which the error appears, and in which version that the error was first corrected. There is no good reason why the same principle cannot be applied to academic writing, with each released revision having its own number, the minor numbers (tenths) indicating minor revisions, and the major numbers (units) indicating major revisions, with the first generally released version being numbered 1.0, and pre-release versions 0.x. This way, when a work is cited, the revision number can be cited, too, so that the author is free to amend and improve the work without rendering inaccurate or misleading existing citations. Therefore, **if you are citing this work, please do not cite just the title, but also cite the revision number and date specified below.***

**Revision Number:** 1.0

**Revision Date:** 22<sup>nd</sup> of May 2004

### **Acknowledgements**

I should like to thank: Professor John Gardner, whose refutation of my attempted defences of classical utilitarianism in seminars in Oxford in late 2003 inspired me to consider more deeply the fundamentals of the nature of value, which ultimately lead to many of the important ideas behind the first part of this paper; Gail Leckie, discussions with whom at around the same time about the nature of value and evaluation also lead to some of the important ideas behind this theory of the good; Professor Joseph Raz, whose writings and excellent seminars on his theory of authority helped to inspire the compound theory of value outlined in Part VI of this paper; Ben Delo, whose assistance was invaluable in dealing with the more mathematical elements of Part V; and Dwight Newman and Danny Priel, who, by kindly letting me present these ideas at their Jurisprudence Discussion Group, gave me a reason to put all of this in writing in the first place.

## **Introduction**

The nature of, and the connexion, if any, between the right and the good has been one of the most enduring debates among philosophers over the centuries, partly because such a debate is inevitably important<sup>1</sup>, and partly because of the great mysteriousness that seems to envelop the topic: the question, “why are some things good and bad; right and wrong?” seems to many as illusive as, “why is there something rather than nothing?”, or, “what is consciousness?”. Few, it seems, have difficulty in disagreeing with David Hume’s remark that,

*“...the distinction of vice and virtue is not founded merely on the relations of objects, nor is perceived by reason.”*<sup>2</sup>

In a similar vein, many, no doubt, would concur with Moore when he wrote,

*“[The good] cannot be defined, and that is all that I have to say about it”*<sup>3</sup>.

Theories about the good and the right have been almost as diverse as is human imagination, and include (amongst, no doubt, many others) those that fasten the concepts to the notion of the will of an omnipotent creator-god; those that argue that what is right to do is whatever people feel like doing; those who suggest that a “reflective equilibrium” of people’s beliefs about justice determine the right, and that the good follows from that; those that postulate that the greatest happiness of the greatest number is of ultimate good, and that the right thing to do is maximise it; and those who deny the existence of both concepts entirely, claiming that they are both illusory.

Despite the apparent shroud of mystery, however, I will argue that Hume was wrong to claim that the nature of vice and virtue, in other words, value, cannot be discovered by reasoning; that value is a contingent fact of existence, dependant on the prior existence of life, and that, far from being mysterious or non-existent, value is a concept whose nature and application can be arrived at by sound, empirical reasoning.

Perhaps value’s ostensible mystery is derived from the perennial confusion between the two distinct questions: (1) what is value? and (2) what is valuable? Many theories have answered the second question without addressing the first, or assumed that the answers to both questions are necessarily the same. I suggest that such an approach is confused, and fails to elucidate the concept of value clearly. Even Moore, who pointed out such a flaw in other’s theories, committed it himself in his attempt to prove that good is indefinable. This theory addresses each of the two questions separately, and holds that the meaning of the good and the right cannot be understood merely in terms of each other, but must be understood in terms of a specific sort of logical structure into which states of affairs in the world are sometimes, contingently (and, ultimately, by chance) arranged.

I deal first with the methodology of value, arguing in favour of an empirical approach; I then deal with the nature of value itself, that it is derived from life, and what that entails; after that, I examine some specific issues, namely the relationship between different evaluative concepts, and some logical principles that apply to value (including a refutation of value-incommensurability); before finally turning to a theory of the right, and some of the more abstract things, at least, that that entails.

---

1 Indeed, it necessarily defines what importance is.

2 *A Treatise on Human Nature*, Book III, Part I, Section I

3 *Principia Ethica* (2<sup>nd</sup>. ed., Cambridge, 1993), p. 58, ch. 1.

### **Part I – Why an empirical theory?**

An empirical theory of something is a theory which founds its propositions in observations of states of affairs of the world<sup>4</sup>. An empirical theory is true if (1) its observations are accurate, and (2) the conclusions drawn from those observations are drawn logically. An empirical theory of stars, for example, is one that deduces the nature of stars by reasoning from the observations made by astronomers: that they are all moving away from one another because the light that they emit is slightly more red than it otherwise would be; or that different stars are formed of different gasses because the spectrum of light that they emit corresponds to that reflected by different particular gasses observed on Earth. Those theories are true if their observations are accurate and their deductions logical. An empirical theory of value holds that the truth of statements such as, “I must not kill people for my amusement” or, “you ought not set fire to your neighbours’ house because you dislike them” can be derived in essentially the same way: by making deductions from observations of physical states of the universe, and it can be validated or falsified by, and only by, assessing the accuracy of the observations and the logic of the deductions upon which they are founded.

This sort of theory is not compatible with Hume’s popularly accepted assertion<sup>5</sup> that an “ought” cannot be derived from an “is”: indeed, a value empiricist must reject this non-derivationism. Before turning to the reasons for its rejection, two kinds of non-derivationism must be distinguished, only one of which will be rejected. The first is circular non-derivationism: it holds that state of affairs X cannot be valuable<sup>6</sup> merely because state of affairs X obtains in the world. This will not be rejected: if circular non-derivationism were false, then that something is valuable could not mean more than that it exists, and it would thus not be possible to distinguish between valuable, available and antivaluable states of affairs: in short, there would be no such thing as value. The second sort is linear non-derivationism: it holds that state of affairs X cannot be valuable because state of affairs Y obtains in the world. It is this linear non-derivationism that is to be rejected.

Linear non-derivationism must be rejected because it inevitably leads either to scepticism or mysticism. It does so in one of the following two ways:

*(1) one cannot derive any “ought”<sup>7</sup> from any “is”<sup>8</sup>;*

*(2) there are only “is”es;-*

*and therefore*

*(3) there are no oughts;*

(scepticism) or;-

*(1) one cannot derive any “ought” from any “is”;*

*(2) there are “ought”s;-*

*and therefore*

*(3) there are things that are not “is”es<sup>9</sup>*

(mysticism).

---

4 I use “world” and “universe” interchangeably in this paper.

5 Made in *Treatise*, supra.

6 And thus a state of affairs which ought to be brought about: the connexion between these two propositions is explored in Part IV below.

7 A true evaluative proposition

8 A true proposition about a physical state of affairs in the world

9 Or, more eloquently, there are things that exist “outside” the physical world

The only way, therefore, of holding both (1) that there exists nothing beyond the natural world<sup>10</sup>; and (2) that there are, nonetheless, truths about value or morality<sup>11</sup>, without inconsistency, is to reject the first proposition, that of linear non-derivationism. By doing this, both of the second propositions above can be accepted consistently with each other.

Once it is accepted that one state of affairs can only be valuable as a result of another physical state of affairs in the world, an empirical theory of value becomes most attractive. Human experience has consistently demonstrated that by far the most reliable way of discovering the truth about physical states of affairs in the world, and what is entailed by them, is by observation and deduction, and indeed by that particular sort of combination of observation and deduction called “science”. It is thus as much a question for science to answer as to why it is true that people ought to do or not to do certain things as it is why it is true that there are people. It turns out, as one might expect, that the answers to both of those questions are closely connected (although not, I must emphasise, the same), but more of that in Part II, below.

Indeed, that the nature of value can only be discovered by scientific method was recognised as early as 1863, when John Stuart Mill wrote,

*“...In ethics, as in all other branches of scientific study, the consilience of the results of both these processes [i.e. (1) empirical generalisations from the observed results of conduct; and (2) deduction from the laws of life<sup>12</sup> and the conditions of existence], each corroborating and verifying the other, is requisite to give to any general proposition the kind degree of evidence which constitutes scientific proof.”<sup>13</sup>*

For this reason too, attempts by moral philosophers to derive value ultimately from intuition, a “sense” of morality (as distinct from the physical senses), or anything of the sort, must be rejected as being unreliable, for the same reason as a theory of stars<sup>14</sup> based on what people feel about stars or on what would make the best story if it were true must be rejected. There is no more reason to believe that what we feel that we ought to do<sup>15</sup> truly represents what we ought in fact do than to believe that what we feel that stars are truly represents the nature of stars. Mill recognised this when he wrote,

*“that a feeling is bestowed on us by Nature, does not necessarily legitimate all its promptings.”<sup>16</sup>*

The enquiry into value, then, is to how it can be that particular states of affairs make other states of affairs valuable, which states of affairs in the world make things valuable, which sorts of states of affairs are thereby made valuable, and what exactly the consequences of a state of affairs being valuable are. In other words: what is the structure of value, and what are its contents?

---

10 An empirical theory of value therefore rests on a naturalist (realist/materialist/physicalist) notion of the world; since this is not a philosophy that I have space to defend here (but one to which I do subscribe and am prepared to defend vigorously elsewhere), I will proceed on the basis that realism is true, whilst acknowledging here that it is a subject of lively debate in the philosophical community; however, it is useful to split any criticism of my theory into (1) that which rejects this theory because it rejects physical realism, or (2) that which rejects this theory despite accepting it. However, this theory, if successful, *will*, of course, refute arguments against naturalism (etc.) that rest on the proposition that value/morality (etc.) cannot be derived from physical reality.

11 To the impatient sceptic: I argue for the proposition that there are indeed such truths below.

12 There is, of course, an ambiguity as to how “the laws of life” are arrived at: whether by empirical or non-empirical means. However, given the comparison of ethics to science, I suggest that the more faithful interpretation is one that regards these two sorts of basis as different sorts of empirical observation and deduction.

13 *Utilitarianism*, ch. 5 (footnote: reply to Herbert Spencer)

14 Or, indeed, the mind: the psychology is as much of a science as astronomy.

15 Or anything derived from what we feel that we ought to do, without, at least, a reason resting on deductions from observations as to why this feeling should be a starting point in any particular way: it is for this reason that constructivist theories are also to be rejected.

16 *Utilitarianism*, ch. 5

## **Part II – Value and Life**

In this part, I will argue that the concept<sup>17</sup> of value is necessarily derived from – and only from – the concept of life, or, more precisely, that value, as a property of a state of affairs, is only a property of that state of affairs relative to some living entity or entities, or an entity or entities created thereby. Indeed, it is a defining property of life that it is the only sort of process that can create entities for which different states of affairs can be differentially valuable. This is a property of life which is responsible for all of the unique characteristics of living things, in a way which I shall describe below. In short, life and value are inexorably intertwined as conceptual entities. The impact of value on the nature of life has been extensively explored by biologists, and has led to the most fundamental and important understandings about the nature of life, about which more below. The impact of life on value, however, has been sadly neglected by (many) moral philosophers, and it is this deficiency that I hope here to remedy.

Firstly, briefly to clarify what it is for a state of affairs to have a property only in relation to another state of affairs. I hope that this paragraph – if it is needed at all – is needed only to pre-empt any semantic misunderstandings of the first sentence of this part. An elephant is big: that is a state of affairs in the world in the form of a property of elephants; but its bigness is relative to other things that are not elephants: humans, for example. An elephant is decidedly not big compared to planets, even relatively small planets. An elephant has the property of being big relative to humans (and other things of that size). It does not have the property of being big relative to other elephants, or to planets. In the same sense, a state of affairs has the property of being valuable, or not, relative only to living entities, or entities created by living entities. Of course, at the most fundamental level, every property of every state of affairs in the universe exists relative to other states of affairs in the universe: this is the theory of relativity at its most banal. There is no extra-special degree of relativism with value: every property of every state of affairs exists (in so far as it does exist) in relation to some other state of affairs, or property of that state of affairs: for value, it is life, or, indeed, as I shall explain below, particular living entities.

The following expository thought experiment should demonstrate the necessity of the connexion between life and value. Imagine a universe totally devoid, now and for ever in the future, of any conceivable form of life. Could any one state of affairs in that lifeless universe be any more or less valuable than any other? However theories of value differ in methodology and content, they all converge in answering that question in the negative. For intuitionists, there are no organisms that can have intuitions; for constructivists, there are no constructors; for utilitarians, there is no pleasure, pain, nor desire; for perfectionists, no creatures capable of flourishing; and so on. There does not seem much dispute, then, that there is *some* necessary connexion between life and value, which by itself would demonstrate that value, unlike, say, mathematics, is a contingent, rather than a necessary feature of the universe. Why it is true – and what this connexion entails – I explain below.

Value and life share in common a specific sort of characteristic not shared by any other entity in the universe: decisionmaking. By “decisionmaking”, I do not just mean the conscious taking of decisions, although conscious decisionmaking is invariably a subset of the broader concept: I mean a logical structure of causality that is unique – and fundamentally important – to life. Life is unique in tending to bring about states of affairs because of the nature of the states of affairs so brought about: value is those properties of those states of affairs that cause it to be the case that entities tend to bring them about. Decisionmaking is the act of bringing about valuable, as opposed to invaluable or antivaluable, states of affairs, in consequence of their value. A more comprehensive explanation is in order.

What makes it the case that an entity is alive? Suppose that, some distant year in the future, exobiologists from earth ventured forth to search for life in other galaxies: how would they know

---

17 In the sense of a universal entity with common properties, not necessarily an idea held by a sentient being.

whether they had found it? They could perform chemical tests on anything that they find to see whether it contained the sorts of chemicals that life on earth contains, but there would be no way of knowing whether life in distant galaxies could be based on some other sort of chemistry. They could look to see whether there was anything that was moving around, but how would they tell the flowing rivers, the scudding clouds and the whirling sandstorms apart from the movement of living creatures<sup>18</sup>? Alternatively, they could look for *order*: apparently purposeful movement, or a regularity of complexity that would be extremely unlikely to occur by chance. But order is more than mere regularity: the layers of sedimentary rocks; snowflakes; and the ripples of sand on a beach at low tide all have ostensibly improbable regularity, and yet all have non-organic causes. Order, as distinct from mere regularity, is an essentially *evaluative* concept: an ordered state of affairs is a state of affairs that is regular or uniform in a particular way for a *purpose* or a *reason* or, minimally, because being that way serves some *value*. A test for life that rests on an evaluative criterion, of course, begs rather than answers the question as to the relationship between value and life; additionally, it does not answer our exobiologists' question, since they will need to identify what value is in order to distinguish between order and mere regularity. Indeed, the need to resort to evaluation in the identification of life in this way is a clue to the close nexus between life and value.

Of course, the only sure (and non-evaluative) test for life, and that which ultimately defines life, is self-replication. If the exobiologists found an entity of any sort that reproduces itself<sup>19</sup> over multiple generations, then they could be as sure as they were ever going to be that they had found life. Deductions from observations have concluded that this self-replicative property has the following effect: of those entities that reproduce themselves, only those that do so consistently over many generations will be left after the passage of some time: as Dawkins<sup>20</sup> explains, those replicative entities that tend to be left after some time has elapsed are those that have the three properties, *stability*, *fecundity* and *copying fidelity*. As the number of replicators increase, so the environment for replication changes, with the consequence that entities with different (and usually more complex) sets of characteristics than before are stable, fecund and make faithful copies of themselves in the new environment. Those that no longer have those properties (relative to the new environment) in sufficient quantity eventually cease to exist. This is a process which, given the limited resources of the environment, leads to spiralling complexity, and, ultimately, the sort of inordinately complex life forms found on earth. That, with inordinate brevity, is the familiar theory of evolution, and so far is nothing new.

Notice, however, one striking feature of life: unlike any other entity in the universe, life is the only sort of thing that behaves<sup>21</sup> in any particular way in consequence of the effect that that sort of behaviour tends to have had in the past, in causing the entity that behaved in that way to possess sufficient stability, fecundity and copying fidelity (which combination I shall call "replicativity" for short) that its copies go on to replicate themselves and continue to populate the world. In other words, living things are the only sorts of things that behave in a given way as a result of the likely *future consequences* of behaving in that way (that circumstances are sufficiently consistent that past tendencies provide a reasonably accurate guide to future probabilities is a necessary condition for the existence of life on any evolutionary account). As Dennett wrote,

*"In the beginning<sup>22</sup>, there were no reasons; there were only causes. Nothing had a purpose,*

---

18 Of course, plants, a very important category of living creatures, do not move very fast at all, although evidence of their past, slow, movement would not be too hard to find.

19 Of course, by "itself", I mean a complex pattern of characteristics, not a mere property of matter such as heat or an exothermic chemical reaction: fire, for instance, is not life, since it does not pass any specific characteristic onto the next "generation".

20 *The Selfish Gene* (2<sup>nd</sup> ed.), ch. 3 (Oxford, 1989)

21 In the broadest sense of the word; in the sense in which the Earth "behaves" by orbiting the sun, or the seas behave by tidal fluctuation.

22 By which Dennett means, before there was life.

*nothing had so much as a function; there was no teleology in the world at all*"<sup>23</sup>.

Nebulae do not form themselves into stars because stars tend to throw large quantities of electromagnetic radiation into the surrounding space; nor do meteorites strike our moon's surface because meteorite strikes tend to produce craters. However, rabbits *do* run away from foxes because doing so tends to mean that they will not get eaten; and squirrels *do* bury nuts because burying nuts tends to cause them to be more likely to have enough food in the winter. Self-replication, then, in sufficiently stable conditions, in and of itself, produces entities whose nature and behaviour is determined at least partly by the consequences that that nature and behaviour tends to have. No cause that is not part of or created by a self-replicating process can ever be partly constituted by the likely effect of that cause: it is only the feedback made possible by differential survival of imperfect copies that enables probable effects to influence causes. It is this characteristic that I call decisionmaking, and it is this characteristic that fundamentally defines both value and life, as well as the close and necessary nexus between them.

I have called that characteristic of states of affairs that tend to be caused by the entities that cause them, at least in part, because that characteristic is part of the effect, "value"; but this is not a special, technical definition of value, applicable only to evolutionary theory: it is the very essence of what it is for anything to be valuable, on any meaningful account. In other words, value necessarily depends on the concept of decisionmaking, and therefore life, for its meaning.

Indeed, all theories that purport to be theories of value have inevitably incorporated this structure, albeit not always expressly. Teleological theories, those that hold that the good is prior to the right, entail that a characteristic of a state of affairs (goodness) determines whether it is right to bring about those states of affairs, and therefore that a characteristic of a potential future state of affairs can constitute a reason to bring it about. Since a reason is a form of cause, a teleological theory necessarily rests on the logical structure outlined above, that characteristic of a state of affairs can cause it to be brought about.

Deontological theories of value, those that hold that the right is prior to the good, have this essential structure: all of the actions people can choose to perform have a characteristic of being differentially right. There may be a sliding scale of rightness that descends well into the negative, or there may be only a few (at minimum, two) categories into which all actions fall (right and wrong). The more actions that are performed that have the property of rightness (and, where relevant, the more rightness that those actions have), the more value that there is in the world, and the converse also holds.

This structure, too, cannot avoid but employ the concept of decisionmaking in the way that I have described it: if an action's being right is a reason to perform it, then something in the nature of the action (its rightness, or whatever constitutes that) can be part of the cause of its being performed. Since the cause of the action being performed must pre-date the performance of the action, this notion, too, rests on the importation of the effect-tendency into the cause. Indeed, even in respect of a theory that held that *decisions*, rather than *actions* were the bases of right, it would still be the case that, if the theory had any weight at all, the decisions would themselves be (at least in part) consequences of their rightness, which rests on the notion decisionmaking just as much as any teleological theory. Without the feedback that can be created only, ultimately, by systems of self-replication, no property of an action or decision can cause it to be instanced or decided.

It can now be seen that the concept of decisionmaking explains the relativity of value: a state of affairs can only be valuable relative to some particular decisionmaking system. It is entirely possible for there to be more than one system of this sort in the world, for which the same state of affairs is differentially valuable. Indeed, that there are many such conflicting systems is a vital constituent of



any theory of evolution<sup>24</sup>.

No doubt this theory will raise the criticism that it commits what Moore<sup>25</sup> called the “naturalistic fallacy”: it is fallacious, Moore writes, to conceive of value as a complex of natural properties. This theory sets out to reject the notion that ethical naturalism is fallacious. In support of his argument, Moore wrote,

*“The hypothesis that disagreement about the meaning of good is disagreement with regard to the correct analysis of a given whole, may most plainly be seen to be incorrect by consideration of the fact that, whatever definition be offered, it may always be asked, with significance, of the complex so defined whether it itself is good”<sup>26</sup>.*

Moore suggested that if, for example, one were to suggest that good is what people desire to desire, then the statement that what people desire to desire is good means more than that what people desire to desire is what people desire to desire, or that the word “good” is used always and necessarily to mean that which people desire to desire. In making this point, however, Moore commits the very logical flaw which he condemns in others<sup>27</sup>: that is conflating the question of what value is with the question of what is valuable. If what people desire to desire (or anything else) is offered as an understanding of the inherent nature of value, it necessarily cannot be meaningful to enquire whether it is valuable, any more than it can be meaningful to enquire whether any proposed definition of height really is tall. If what people desire to desire (etc.) is offered as proposition about what is ultimately valuable, then, of course, it is meaningful to enquire whether what people desire to desire (etc.) is really valuable, in the same sense as it is meaningful to enquire whether any significant proposition is true. However, if Moore meant no more than that a proposition about what is valuable cannot, by itself, serve as a proposition about the nature of value<sup>28</sup>, then he has not established that propositions about the nature of value cannot, as he claims, be derived by deductions from observations, since it is perfectly possible to have a proposition about the nature of value that does not simultaneously claim to be a proposition about that which is valuable: the proposition presented in this theory about the logical structure of value is one such proposition, and Moore offers no argument about why propositions of that nature cannot be true.

So, we have seen that life is the only kind of entity in the universe capable of creating systems of decisionmaking; that decisionmaking is an essential component of value; and that value is relative to particular decisionmaking systems. But, as far as human minds are concerned at least, what property of states of affairs is valuable? The answer, as ever, lies in deductions from observation, and, in this case, they lead to the concept of orders of value.

---

24 As Dawkins clearly explains in *The Selfish Gene* (supra).

25 In *Principia Ethica*, supra ch. 1.

26 Ibid, p. 67, ch. 1.

27 He wrote, “*We may... mean to ask not what thing or things are good, but how ‘good’ is to be defined*”, p. 57, ch. 1

28 And with this I should agree.

### **Part III – Decisionmaking, orders of value, and pleasure**

Dawkins famously wrote in the opening chapter of *The Selfish Gene*,

*“I am not advocating a morality based on evolution. I am saying how things have evolved. I am not saying how we humans morally ought to behave.... If you wish to extract a moral from it, read it as a warning. Be warned that if you wish, as I do, to build a society in which individuals cooperate generously and unselfishly towards a common good, you can expect little help from biological nature. Let us try to teach generosity and altruism, because we are born selfish. Let us understand what our own selfish genes are up to, because we may then at least have a chance to upset their designs, something that no other species has ever aspired to do.”*

Is this theory, then, not precisely what Dawkins (perhaps aptly, one might think) warned against? Is *An Empirical Theory of Value* not just that: a “morality based on evolution”? Well, yes and no. Yes, in the sense that it holds that there can be no morality, no virtue, no goodness, no righteousness; no value, without a process of natural selection and the systems of decisionmaking that only it can create, and that all evaluative concepts are created always by, and only by, such a process. No, in that this theory, as this part explains, does not hold that the values that the process creates are necessarily confined to replicativity (although the status of replicativity as a value is a prerequisite to anything else’s status as a value). In particular, it emphatically does not hold that the sort of value in which humans are naturally (and reasonably) most interested – that which is valuable for conscious human minds – is that replicativity. On the contrary, as I am about to explain, that which is valuable for human minds is an entirely distinct second-order sort of value, independent of its first-order genator, replicativity, which is valuable for human genes. But, like all value, it is created by, and can only be fully explained by reference to, that very replicativity. In short, I will explain not just why this theory does not ignore Dawkins’ warning: I will explain how his warning can make any sense at all.

To summarise the position so far: there can be no value without the sort of feedback caused by the phenomenon of evolving life that permits the logical structure of decisionmaking to operate upon causes. The sort of value that any system of evolving life must create in order to evolve is replicativity, and the existence of replicativity as a value is an inherent part of the logical structure of evolution. Value is always relative to some entity, and, in the case of replicativity, it is relative to particular replicators (genes<sup>29</sup>). Any given state of affairs can be – and often is – differentially valuable to different genes, and the effect of this conflict is an important aspect of evolution.

The point of this part, however, is this: given that the process of evolution, with its feedback-induced predicative decisionmaking, exists, that process can create subsystems, with their own feedback and predicative decisionmaking, and with their own, wholly independent values. I shall call these values “second-order” values, and replicativity the “first-order” value. The only reason that any such system could come to exist is because it tends to serve a higher order value<sup>30</sup> better than some existing or readily-evolvable alternative<sup>31</sup>, but that does not make the second-order values any less independent.

Second-order values have the same logical structure as the first-order value, and are identical except that their content is not replicativity. If a causative event occurs in consequence of the tendency in the past for causative events with properties possessed by that event to bring about

---

29 Dawkins (supra) suggests that genes be individuated by their individual phenotypes – the specific impact that they have upon the world.

30 Either the first-order value, or a value that serves the first-order value, or a value that serves a value that serves..., etc.

31 The exact mechanism for the evolution of second-order values is, mercifully, the task of the evolutionary biologist, and not the moral philosopher to explain.

states of affairs with particular properties, then, where those properties are not themselves replicativity, they constitute a second-order value. Entities that have the property behaving in a particular way as a result of the tendency of that behaviour to bring about states of affairs with properties that are not themselves replicativity can only be created if entities with such tendencies tend to serve that first-order value more than entities with other tendencies.

Second-order values, in order to be independent, require independent decisionmaking systems: genes as a decisionmaking system will always operate so as to further the first-order value of their replication. What does it take for something to be a system of decisionmaking that is, in the requisite sense, independent from the decisionmaking system of gene-selection? What is required is independent feedback. An example will perhaps best illustrate the notion of independence. In order to survive and multiply, a gene<sup>32</sup> might (for example) have one of two sorts of effects on its organism<sup>33</sup>: I shall call the first “the robot”, and the second “the thinker”<sup>34</sup>. Robot genes might cause the organism in which they are resident to perform act A in condition X, act B in condition Y, and act C in condition Z, because performing those acts in those conditions tends to increase the replicativity of the gene that causes that behaviour. The robot is not an independent decisionmaking system, since the particular behaviour, though conditional, is chosen directly by the mechanism of gene-selection: if the environment changed so that, now, in condition X, behaviour A<sub>1</sub>, in fact, was more conducive to survival, the only way in which a creature of that sort could come to perform act A<sub>1</sub> on condition X (consistently) would be if a specific gene for doing A<sub>1</sub> on condition X came to be prevalent as a result of it having an advantage in the gene pool over its if X then A rival<sup>35</sup>. The only sort of feedback, in other words, that determines the robot’s behaviour, is that of differential survival, which is, of course, not independent of replicativity.

Genes for thinkers, meanwhile, might cause their host creature to act in whatever way as would result in the existence of a particular state of affairs inside that creature’s body<sup>36</sup> (V), through a complex mechanism of retention of information about the consequences on that state of affairs of past actions and control mechanisms for present actions: this would indeed be an independent system, with independent feedback. Unlike the robot, the thinker could adapt, without any change in its genes, to different environments where different behaviours are conducive to replicativity<sup>37</sup>. If performing act A in condition X most increases V, then the creature will tend to perform act A in condition X. If, conversely, performing act A<sub>1</sub> on condition X would most increase V, then the creature would tend to do that instead, without any change in its genes<sup>38</sup>. Significantly, however, even if the environment changed so that, whilst performing act B on condition Y was still most conducive to replicativity, performing act B<sub>1</sub> would now most increase V, then the creature would tend to perform act B<sub>1</sub>. Notice that “robot” genes could never cause any entity consistently to perform act B<sub>1</sub> on condition Y in these circumstances. Hence, the feedback (to V) that determines behaviour is, in a relevant sense, independent of the feedback of differential survival. Of course, the feedback of differential survival may still operate on such creatures: a species in which the thinker gene is very common might become extinct if its internal feedback was too far at odds with that which is conducive to its replicativity; or a gene in that species’ gene pool that causes the (robotic) if Y then B tendency, partially over-riding the “thinking” system, might become prevalent, but if the if Y then B<sub>1</sub> tendency is not so detrimental to replicativity so as to make the species become extinct, and if no if Y then B gene can become prevalent without doing more harm than good to

---

32 Or set of genes – the individuation of genes does not much matter for this explanation

33 Its “phenotype”.

34 This is, of course, a hypothetical and highly simplified example.

35 In theory, a robotic species could evolve into a non-robotic species, of course, but that is not the point here.

36 Whatever combination of sufficient warmth, food, water, satisfaction of sex-drive, etc. is evolutionary optimal

37 This may be an explanation of how more intelligent forms of life evolved, although that is not the purpose of this observation.

38 Provided, of course, that it could work out how.

replicativity, then the if Y then B<sub>1</sub> tendency would nonetheless prevail.

In the thinkers' second-order value system, V represents the independent, second-order value. V is valuable for the independent decisionmaking system that the thinkers' genes have created, even when it is valuable or antivaluable for those genes. It is my contention that human minds, like the hypothetical thinkers, are independent decisionmaking systems, and that pleasure is their value.

That pleasure is of ultimate value to human minds is an idea as familiar as it is old: as early as 1789, Jeremy Bentham wrote,

*“Nature has placed mankind under the governance of two sovereign masters, pain<sup>39</sup> and pleasure. It is for them alone to point out what we ought to do, as well as to determine what we shall do. On the one hand the standard of right and wrong, on the other the chain of causes and effects, are fastened to their throne. They govern us in all we do, in all we say, in all we think: every effort we can make to throw off our subjection, will serve but to demonstrate and confirm it”<sup>40</sup>.*

It is interesting to notice that, even in 1789, Bentham had realised the necessary connexion between causality and value, and also (as entailed thereby) rejected linear non-derivationism, holding that a state of affairs is valuable in consequence of a physical state of affairs in the world (“nature”), derived, one can only imagine, from deductions from observations. Alas, Bentham did not spell out his methodological reasoning in this respect more fully. Mill, however, did:

*“Questions about ends are... questions what things are desirable. The utilitarian doctrine is, that happiness is desirable, and the only thing desirable, as an end; all other things being only desirable as means to that end.... / The only proof capable of being given that an object is visible, is that people actually see it. The only proof that a sound is audible, is that people hear it: and so of the other sources of our experience. In like manner, I apprehend, the sole evidence it is possible to produce that anything is desirable, is that people do actually desire it.”<sup>41</sup>*

Mill was clear, at least, as to the empirical nature of his theory of value. But what exactly is pleasure; and why should it, as opposed to anything else, constitute the ultimate value for the decisionmaking system of human minds? The theory that pleasure is of ultimate value to humans has had many detractors: Rawls, for example, writes,

*“Surely, the preference for a certain attribute or feeling or sensation above all else is as unbalanced and inhuman as an overriding desire to maximise one's power over others or one's material wealth;”<sup>42</sup>*

and Moore wrote,

*“There is no meaning in saying that pleasure is good unless good is something different from pleasure”<sup>43</sup>.*

---

39 Bentham and Mill both wrote of pain and pleasure separately, but I think that it is more productive to conceive of pleasure as a continuum that can be both positive and negative. I suspect that Bentham and Mill meant no more than the negative (as opposed to the absence) of pleasure when they wrote of pain, but that may be subtly different to the common understanding of the notion of pain: as Gardner once pointed out, masochists positively enjoy “pain”, in one sense of the word, yet I suspect that Bentham and Mill would regard masochists as receiving pleasure when they experienced pain in that sense. Sidgwick was clear in his understanding of pain, when he wrote, “...pain must be reckoned as the negative quantity of pleasure, to be balanced against and subtracted from the positive in estimating happiness on the whole” (*Methods of Ethics*, Book II, ch. 2, §1).

40 Introduction to the Principles of Morals and Legislation, ch. 1 (1). Emphasis from the original.

41 *Utilitarianism*, ch. 4

42 *A Theory of Justice* (2<sup>nd</sup>. ed, 1999), p. 488

43 *Principia Ethica* (supra), p. 66, ch. 1.

In one sense, Moore's criticism is no criticism at all, since this theory<sup>44</sup> does not postulate that pleasure is *the same thing as* value, but that pleasure is ultimately valuable for conscious human minds. However, Moore seems to see his "naturalistic fallacy" objection as supporting a criticism of theories that hold that pleasure is of ultimate value as well as theories<sup>45</sup> that hold that pleasure is, ultimately, value. Nonetheless, he seemed to believe that all of those theories that had so far been presented that made a claim as to that which is of ultimate value did so merely by contending that the word "value" (or "good") meant no more than whatever it was that they claimed was of ultimate value<sup>46</sup>. Since the claim made here is that pleasure is of ultimate value to human conscious minds because it fits precisely the logical structure that I have already established constitutes the inherent nature of value (as explained more fully below), not because the word "value" merely means pleasure, Moore's criticism is not a criticism of this theory. It is as significant and meaningful to state that pleasure is of ultimate value for human conscious minds as it is to state that our sun is a star, and no less true for it.

Turning to Rawls' criticism, his use of evaluative concepts ("unbalanced and inhuman") to assess the truth of a theory of value presupposes rather than establishes the truth of a rival theory of value against which the former can be assessed. One can no more evaluate a theory of value than ascribe financial worth to the concept of money, or measure the size of bigness. As explained in Part I, a theory of value stands or falls only on the extent to which its observations of states of affairs in the world are accurate, and its deductions from them logical; to be meaningful, a theory of value must explain its object in non-evaluative terms.

In order to test whether pleasure is the value of the second-order decisionmaking system of human conscious minds, the enquiry must be directed at whether the concept of pleasure matches the logical structure of value presented in Part II above, and whether it is a concept that applies to the decisionmaking system of conscious minds. As explained above, value is those properties of those states of affairs that cause it to be the case that entities tend to bring them about. Pleasure is the value of human conscious thought if it is the property of states of affairs that causes it to be the case that human minds tend to cause them. If pleasure is taken to mean a property of conscious experiences<sup>47</sup> that, all other things being equal, of itself tends to cause the mind having those experiences to take such decisions as to cause it to tend to bring about the circumstances that cause those experiences more, rather than less<sup>48</sup>, then pleasure is necessarily that which is valuable for conscious minds.

But is that the true nature of pleasure? It is but a truism that pleasure consists in pleasant experiences; and what else can the property of pleasantness, described in the non-evaluative terms that any theory of value must use, be but that it is that property of experiences that, all other things being equal, tends to cause the mind having those experiences to bring about the circumstances that cause it to have them more often than it otherwise would? It is, of course, entirely possible that there are people who understand "pleasure" in some narrower sense, and perhaps the notion of that narrower form of pleasure as the ultimate value has been what has dissuaded some from endorsing a theory of (human) value based solely on pleasure. However, it is clear that Mill, at least, meant pleasure in the broader sense in which I have used it when he wrote, "*the sole evidence it is possible to produce that anything is desirable, is that people do actually desire it*"<sup>49</sup>; and Bentham's references to causality suggest that he, too, subscribed to this wider notion.

---

44 And arguably Bentham and Mill's, although Moore did not interpret them this way.

45 Which it can be argued that nobody has ever advanced.

46 In particular, he criticised Bentham for this flaw, but it is clear from my quotation of him above, that he recognised the essential link between value and causality, and therefore did not merely seek to define good as pleasure, but provided reasons as to why pleasure is ultimately good.

47 Which are physical states of affairs in the world, just like anything else

48 And for displeasure, less, rather than more

49 *Utilitarianism* (supra).

All that is further needed to demonstrate that the concept of pleasure applies to human minds is evidence that there are certain experiences that, of themselves, have the property of tending to cause those who experience them to want to experience them again, and more of them, and some experiences that have the converse tendency. Any experience of being human, I suggest, contains experience of this structure; the evidence for it is the nature of what it is to be alive and conscious.

A theory of value based on considered judgments, on intuitions, or on anything of the sort, is therefore bound to fail for the simple reason that the very existence of judgments, intuitions or any form of decisionmaking, and their capacity for meaningfulness, is inexplicable without reference to value in the sense in which I have described it. For all of these reasons, therefore, pleasure, in the broader sense of pleasant experiences, is the ultimate value for the human conscious minds that have those experiences.

Of course, human conscious minds are more complex than the hypothetical thinkers. A vast array of different sorts of experiences are pleasant or unpleasant to human minds, and they are pleasant and unpleasant in different degrees. The logical structure of degrees of pleasantness is as follows: an experience (A) is more pleasant than another experience (B) when the nature of experience A causes, all other things being equal, the (conscious<sup>50</sup>) decisionmaking system for which it is pleasant to tend to bring about the circumstances for experiencing A more often than for experiencing B, even when B is more pleasant than some other experience (or the lack of experience) for the same reason. As always, the structure is mirrored and reversed for unpleasantness.

Moreover, humans are not pure “thinkers”; the existence of reflexes demonstrates that human action is determined, at least in part, by “robotic” tendencies<sup>51</sup>, and, arguably at least, instincts, intuitions, gut feelings and other such mental constructs are some sort of hybrid between “thinking” and “robotic” decisionmaking (for example, if X, then doing A will feel pleasant, and not doing A will feel unpleasant, irrespective of the eventual consequences of doing A). If this is the case, the probable result of following instincts, intuitions and gut feelings is that doing so will be more likely to benefit human genes, and less likely to benefit human minds than acting rationally (the pleasantness of the action might well not outweigh the unpleasantness of its consequences, for example). This is important to the question of decisionmaking method, and its relationship with rationality: concepts which are explained further in parts IV and VI, below.

I hope that it is now clear how it is possible to hold (1) that value cannot exist without gene selection, and its first-order value of replicativity, but (2) that, because of the existence of second-order values, what is valuable for human minds is not necessarily the same as what is valuable for human genes<sup>52</sup>. Only the notion of second-order values, indeed, is capable of accounting for how it could be good for any organism’s mind to “upset the designs” of the genes that created it. Why humans ought to act “for the greater good” is explained in Part VII, below.

---

50 “Pleasantness” is perhaps most usefully used to refer to the value of conscious decisionmaking systems.

51 True reflexes have a stimulus-response cycle that entirely bypass the brain, and therefore conscious thought.

52 Of course, one would expect that they would tend to coincide much of the time, as, in fact, they do: what I deny, however, is that an exact correlation is conceptually necessary.

#### **Part IV – The family of evaluative concepts**

An evaluative concept is a concept which depends for its meaning on the logical structure of value that I have described in Part II above. Such concepts divide broadly into two groups: the extended family: those concepts that rely on the notion of value in general; and the immediate family: those concepts that rely on the second-order value as applied to conscious human minds. I shall concentrate on the second for no other reason than that these relate more closely to the issues in which humans are understandably most interested; however, I shall briefly mention members of the extended family first.

Extended family evaluative concepts are wide-ranging, and cover many areas that are not conventionally conceived as evaluative, since they do not play the same rôle in human decisionmaking as do the immediate family concepts. It is not unreasonable, for the most part, to use the word “evaluative” to refer only to those immediate family of concepts, but there may nonetheless still be some uses in making reference to the extended family concepts as evaluative for explaining their relationship with the logical structure. Replicativity is one of those members of the extended family, for the reasons described above, as are all forms of order that serve non-human values. Indeed, any concept that depends for its meaning on a logical structure of which decisionmaking is a part is an evaluative concept of some sort. It is for this reason, of course, that a theory of value, which must be explained in non-evaluative terms in order to be meaningful, cannot be based, ultimately, on decisions or judgments or intuitions, since they are all evaluative concepts, at least in this extended sense:<sup>53</sup> it makes no sense to conceive of decisions or judgments or intuitions without first conceiving of the logical structure of value.

Turning to the immediate family, I shall describe a number of evaluative concepts that are of importance to the fields of moral and political philosophy by delineating their relationship with the logical structure of the second-order value of human conscious minds that I have explained. Some of the explanations are obvious: I include them for completeness. The purpose of this section is not to be a glossary of evaluative terms; rather, it is to demonstrate how all evaluative concepts can be explained (ultimately) in non-evaluative terms according to the logical structure of predicative decisionmaking. For ease of reference, they are presented as a list.

- *The good* refers to the value of a particular decisionmaking system, and is usually, although not necessarily, confined to the decisionmaking systems of human conscious minds.
- *Better than* and *best* refer, in the comparative and superlative, to degrees of goodness. The meaningfulness of these concepts reinforces the quantitatively comparable nature of value, explained above.
- *Perfection* is that (elusive) state of affairs which cannot be changed in any way to make it more valuable, either relative to the thing in respect of which it is perfect, or (even more elusively) in relation to all decisionmaking systems.
- *Imperative* has many distinct meanings, but one particular meaning refers to the nature of a cause in a system of conscious decisionmaking.
- *Preference* is the consciousness of two or more possible outcomes of a set of circumstances (which may or may not be hypothetical or imagined), and the appreciation that one is better than another (and possibly that they are ranked in order of value).
- *Desire* is the conscious appreciation of the possibility of a future valuable state of affairs, coupled with a preference for that state of affairs over other possible states of affairs, and mental constructs that can (or that can be imagined to be able to, at least in some circumstances that can

---

<sup>53</sup> This is why, when stating that value depends on life, it is necessary to describe in non-evaluative terms the logical structure of life itself.

be imagined) bring about that state of affairs.

- Desirability is, as Mill pointed out<sup>54</sup>, no more than that value itself, or possibly imagined value.
- Purposefulness is the consciousness of the possibility of a causal chain of events leading to a state of affairs that is valuable itself, or that increases the likelihood of other valuable states of affairs arising, combined with a desire to bring about such a state of affairs; *purpose* refers to a state of affairs that has been or may in the future be brought about purposely<sup>55</sup>, or to a decision to bring about such a state of affairs.
- Importance refers to the capacity of things to cause states of affairs to change in ways relevant to the value of those states of affairs; the greater the capacity for value-relevant change, the greater the importance.
- A problem is a state of affairs, or a property of a state of affairs, or states of affairs in general, that causes that state of affairs, or other states of affairs, either specifically or in general, to be less valuable than they would be if the problematic state of affairs was different.
- The right refers to those sorts of decisions (causes) that tend to increase (usually human) value; most usually, that option from a range of known options that most increases (or, where applicable, least decreases) human value<sup>56</sup>.
- Ought is the modality of right actions; i.e. one ought to do (decide, think) something where it would be right to do so.
- Justification is a proof, demonstration or reason to believe that an action or decision was or will be, at least given the limited information available to the relevant decisionmaker, right.
- Justice is the principle that no person shall benefit from her or his own wrong, nor suffer detriment as a result of another's wrong: whilst this is an applied principle, resulting from the application of the theory of right (presented in Part VII, below) to contingent facts in the world, and requires too lengthy an exposition and justification to present in detail in a paper about value in the abstract<sup>57</sup>, a brief elucidation of the first part of the principle is presented in Part VIII.
- Morality and ethics are potentially ambiguous concepts: either they mean no more than that which it is right to do (or the tendency to do the right things), or they refer to a subset of the same; in which case it is unclear why any given, as opposed to any other, subset is of interest, and why any kind of subset is needed at all. The most common subset appears to be that which it is right to do in relation to other people.
- Order, as explained in Part III above, is a state of affairs where some regularity or uniformity serves some value.
- Reasoning is a process of conscious minds that uses logic (and, where necessary, mathematics) to determine what to decide so as to reach the most right conclusion.
- Rationality is the state of mind, or attitude, of using or having chosen to use only processes of

---

54 Supra

55 Since purpose is an evaluative concept that requires life for its meaning, it can therefore be seen that it is incoherent to conceive of life itself as having a purpose.

56 "Most increases...": if one holds that the right thing to do is what tends to increase the extent to which states of affairs are valuable, then it would, of course, be inconsistent to hold that the right is anything less than that which causes the greatest increase (compared to other alternatives): if option A is valuable to the amount of 1, and option B valuable to the amount of 2, then choosing option B over A increases value; if option C is valuable to the amount of 3, then choosing option C over option B increases value (and the converse decreases value relative what it could be, and therefore is wrong). Of A, B and C only C is the right option, all other things being equal. This is why the right is constituted by the *maximisation* of the good.

57 Indeed, it probably deserves its own paper.



reasoning ultimately to reach (at least important) conclusions; a rational decision is one that is reached (ultimately) only by a process of reasoning.

- *Reasonableness* is an attribute of a decision or action, that it is as right as one might expect given the limits of human mental capacity in general, and the specific limitations of the circumstances in which the relevant decision was taken<sup>58</sup>. A decision's being not fully rational is the most common, but not the only, cause of it being unreasonable.
- *A reason* is a logical deduction, or set of deductions, that is/are capable of acting as a cause on a decisionmaking system that operates only by a process of reasoning (this does not, of course, mean that reasons cannot also act as causes on systems that sometimes do not operate by a process of reasoning).

The list is, of course, incomplete; however, I hope that I have outlined some of the most important evaluative concepts, and explained their relationship with the logical structure of value, and how that structure is capable of giving them meaning, ultimately, independently of each other.

Some of the concepts relating to reason would best be elucidated by further explanation (as, indeed, promised above); in particular, the relationships between the right, the reasonable, and the rational. Such an explanation is provided as part of a general explanation of decisionmaking method, Part VI.

---

<sup>58</sup> Given the vagaries of language, it may well be that the word has other meanings, too: indeed, the word in its adverbial form is often used to mean no more than "approximately", although, even there, there is some connotation of the approximation being the best that one would expect of rational decisionmakers in the circumstances.

### **Part V – Value vagueness and the incoherency of incommensurability**

It has often been claimed<sup>59</sup> that value unitarist theories, like this one, and those advanced by the classical utilitarians such as Bentham and Mill, are flawed because there exist a plurality of incommensurable values. I will suggest that this notion misconstrues the nature of value and decisionmaking systems, but first a clarification: my theory is pluralist in one way and unitarist in another. On the one hand, it is pluralist in that it asserts that there can be many different and conflicting value systems in the world. On the other hand, it is unitarist in claiming that there can be only one, ultimate sort of value for each type of independent decisionmaking system, including the decisionmaking system of human conscious minds.

The truth of a value unitarist theory entails the falsity of value-incommensurability, and the truth of value incommensurability entails the falsity of value unitarist theories. Value-incommensurability makes the following claims: a plurality of things are valuable for humans. They are so different from each other that, in many cases at least, they cannot be weighed against each other, or aggregated into a combined total of value. Although making an exclusive decision as to which of two or more differently valuable and incommensurable options to pursue can be important as to which of the range is selected, the decision cannot be taken by reference to which is more valuable, since the value of each is incomparable with the value of each other. Raz writes,

*“A and B are incommensurate if it is neither true that one is better than the other, nor true that they are of equal value”*<sup>60</sup>.

In addition, Raz suggests that option C might be better than option A, but not better than option B, even when neither option A nor option B are better than each other. As to the reason for incommensurability, Raz writes,

*“[Whilst] it is true... that when we express a judgment about the value of options we strive to identify what is true independently of our valuation, but the ranking which determines the relative value of options is not a way of getting at some deeper truth, it constitutes the value of those options.”*<sup>61</sup>

He also writes, in relation to certain sorts of choices,

*“The significance of a choice turns out to have little or nothing to do with the fact that one option is better than the other”*<sup>62</sup>

Thus, the issues are: (1) whether value conforms to the Schröder-Bernstein theorem<sup>63</sup>; (2) whether value possesses the property of transitivity<sup>64</sup>, and (3) whether a decision can precede value, or whether value is necessarily prior to decisionmaking. The logical structure entails an affirmative answer to the first two questions, and also entails that value necessarily precedes decisions (as distinct from non-decision causes). As to the first two issues, since, as explained in Part II, degrees of value of states of affairs differ from each other in the extent to which they cause the decisionmaking systems for which they are valuable, all other things being equal, to tend to bring them about, as distinct from other states of affairs, it cannot fail to be the case that, for any given state of affairs, all other things being equal, a decisionmaking system tends to bring it about more, less, or the same number of times compared to other given state of affairs. For each decisionmaking system, therefore, it cannot fail to be the case that any given state of affairs is more valuable, less valuable, or of equal value compared with any other given state of affairs: this accounts both for the

---

59 For example, by Raz in ch. 13 of *The Morality of Freedom* (Oxford, 1986)

60 Ibid, p. 323

61 Ibid, p. 327

62 Ibid, p. 333.

63 That, if  $\neg A \diamond B$ , then  $A=B$

64 That, if  $A \leq B$  and  $B \leq C$  then  $A \leq C$

transitivity of value, and its conformity to the Schröder-Bernstein theorem. As Mill pointed out,

*“...the truths of arithmetic are applicable to the valuation of happiness, as of all other measurable quantities.”*<sup>65</sup>

As to the third issue, the assertion that a decision can precede value fails to account for the rôle of decisionmaking in value, and value in decisionmaking. As explained in Part II above, a decision is a specific sort of causative process or event<sup>66</sup> that is distinguished from other causative processes or events by being influenced in a particular way by the likely effects of the cause; the features of the effect that influence the cause being value. Thus, whilst it is necessarily the case that the features of a state of affairs that make it valuable can only be valuable as a result of being caused by a process or event that is in some way connected to the decisionmaking system for which they are valuable, it is also necessarily the case that that cause is not a decision unless it was partly constituted by the fact that a cause of that nature is likely to procure a more valuable state of affairs than another cause. So, whilst it is perfectly possible for causative processes to make things valuable by causing them, it cannot be the case that *decisions* make things valuable by having decided them, since a decision cannot exist except in relation to value existing prior to that decision. A theory which does not recognise this is incapable of distinguishing between decisions and other sorts of causes.

Incidentally, it is not necessarily true that all causes that originate in humans, or even from conscious minds, are decisions: it is perfectly possible that some incidental feature of any biological process, even conscious thought, has causative effects that are wholly unrelated to value at all, either typically, or as a result of some malfunction. It is also perfectly possible for such non-decision causes to influence which states of affairs are valuable, or the tendency of valuable states of affairs to be brought about. After all, value is those properties of those states of affairs that, all things being equal, tend to cause the decisionmaking systems for which they are valuable to bring them about more often than they otherwise would; all other things are very often not equal, and incidental causes of biological processes constitute an important category of such inequality.

However, it cannot be the case that, if one of two or more options is no more or less valuable than the other(s), it is important which of the two is chosen. As explained in Part VI below, the method by which the decision is taken may be important, and that method might, in turn, entail one, rather than another outcome, but, all other things being equal, a decision between two options can only be important to the extent that one option is better than the other. As Raz himself notes,

*“Every judgment of importance is evaluative;”*<sup>67</sup>

it cannot, therefore, be the case that importance is not determined by the potential for value-differential in whatever it is that is important.

Incommensurability is usually claimed to be a proof of the existence of a plurality of values: that not all things that are valuable (for any given decisionmaking system) are reducible to one, single, ultimate value. Since, as described above, incommensurability fails, does that entail that value pluralism must also fail; even if there cannot be multiple incommensurable values, can there nonetheless not be a plurality of values without incommensurability<sup>68</sup>? Whilst it is true that different features of states of affairs can have the property of being valuable relative to one decisionmaking system<sup>69</sup>, given the transitivity of value explained above, it is necessarily the case that such features are valuable in a single sense<sup>70</sup> that enables comparison of the degree of value in any given state of

<sup>65</sup> *Utilitarianism*, ch. 5 (footnote: reply to Herbert Spencer)

<sup>66</sup> Necessarily including, but not limited to, those of conscious minds.

<sup>67</sup> In *“Authority, Law and Morality”* (*Ethics in the Public Domain*, Oxford, 1994); and see also my explanation of importance, in Part IV above.

<sup>68</sup> It seems that Moore was a value-pluralist without subscribing to incommensurability.

<sup>69</sup> Indeed, this must be true for it to be meaningful to conceive of different kinds of pleasure, for example.

<sup>70</sup> As in the example of pleasure, however different that differing forms of pleasure are, they necessarily have in

affairs with anything else that is valuable for that decisionmaking system. Indeed, where a plurality of things are valuable for a decisionmaking system, it is therefore necessarily the case that a particular function of those things, and not those things themselves, is of ultimate value for that system. For example, if, for decisionmaking system X, states of affairs with both the property A and the property B are valuable, the the ultimate value for X is not A or B, or even A and B, but something like A+B or AB.

That the arguments in favour of incommensurability have superficial appeal, however, does point to an important feature of value; somebody reading an account of incommensurability may very well be persuaded of it because he or she cannot think how one could possibly conclude whether, for example, a walk in the park is better than a nice cup of tea and a sit down<sup>71</sup>, yet might imagine that a cup of tea and a sit down with a biscuit could be better than one without<sup>72</sup>. Since, as explained above, the transitivity of value makes it necessarily the case that the tea without the biscuit, the tea with the biscuit and the walk are all comparable with each other, on the same scale of value, something other than the logical impossibility of comparison must explain the genuine inability of people to be able to reach meaningful conclusions about the difference in value of such things in many cases: that explanation turns out to be the pervasiveness of uncertainty and vagueness.

That uncertainty is a pervasive feature of thought is an idea as uncontroversial as it is old; at least since Socrates declared that the only thing of which he could be certain was his own ignorance, it has been recognised that there is no genuine certainty to be had. It is almost as uncontroversial, save, perhaps, amongst the most hard-line of sceptics<sup>73</sup>, that there are different degrees of uncertainty in human thought. Whilst I cannot be completely certain of what I did yesterday, for instance, I am less certain of what I will do to-morrow, and still less certain of what my second cousin will do to-morrow. Uncertainty is a quantitative concept.

Vagueness is uncertainty as to degree or amount. It is possible, therefore, for it to be the case that  $A > B > C$ , but, because of the nature of A, B and C, the margin of error in determining their quantity is such that difference below a certain threshold cannot reliably be detected<sup>74</sup>. Thus, in a comparison between A and B, or between B and C, it might not be possible to make a judgment about which is greater than the other, since the difference between the two may be so small as to be below the threshold that below which the differences are smaller than the margin of error. However, the difference between A and C may be so great as to exceed this threshold, making it possible to determine that  $A > C$ , but simultaneously requiring equivocation on whether  $A > B$  or  $B > C$ , or, indeed,  $A = B$  or  $B = C$ .

It is not only threshold vagueness that can make comparisons between theoretically comparable quantities practically difficult<sup>75</sup>: complexity of comparison may make the human mind unable to process the information with sufficient accuracy to render a judgment that is reliable enough to be meaningful. If A, B and C are theoretically comparable, but, whilst A is very similar to B in many respects, it is very different to C (on indices other than that of comparison), then it might be the case that it is possible to judge that  $A > B$ , and that  $A > C$  but not whether  $B > C$ . To give an example, imagine that a tourist visits two grand castles during a visit to England. Both castles are large and impressive, and both are sufficiently in tact for their rooms, of which there are many, to be mainly complete. The tourist spends a day at each castle, and afterwards is asked which of the two castles is

---

common the property of being pleasant to a particular degree, necessarily comparable, in theory at least, with any other degree of pleasantness.

71 The walk in the park is an example from Raz; the tea is inspired by <http://www.nicecupofteaandasitdown.com/>

72 Again, an example adopted from Raz, who prefers to write of brandy and a good book than a cup of tea and a biscuit: *The Morality of Freedom*, ch. 13.

73 The sort who are inevitably reduced to arguing that their own arguments do not exist

74 There may also be higher-order vagueness as to the exact level of the threshold.

75 And sometimes practically impossible

bigger, in terms of indoor floor-space. It would not be surprising if the tourist were unable to answer this question, even if she was able to pronounce confidently that both castles are bigger than her own house, and that either of the castles would be bigger than they already are if they had a single extra room, or even that, if a large hall was built onto one of the castles, that it would, indeed, be bigger than the other in terms of floor space. None of this means that it is not true that one of the castles, as unmodified, *is* either bigger, or the same size, as the other in terms of indoor floor-space, nor that it is not meaningful to compare the floor-space of any given building with any other.

Human pleasure, I suggest, is not always easy for humans to measure in detail, but it is always theoretically comparable. Like the castle, humans tend to be aware of large fluctuations, and to be able to rank significant differences in pleasure, for example, between being stuck in a traffic jam and winning a race, and yet, practically, have difficulty in ranking the pleasure between, for example, tea-drinking (with or without biscuits) and park-strolling. This is doubtless because the particular amount of pleasure that each activity entails is highly variable, depending on (for example, and amongst no doubt many others), how tired that one is, what the weather is like now, and what it will be like for the next hour, how good that the cup of tea will be, what flavour of biscuit that one has to hand, how restless that one is feeling, whether one will come across one's best friend (or somebody whom one would rather avoid) in the park and so forth: all factors which are extremely difficult, and sometimes practically impossible, to determine. Moreover, memories of just how much pleasure that one had when undertaking any given activity are not necessarily very accurate, yet there is no other means than recourse to them in many cases for making evaluative judgments of that sort. Furthermore, pleasure is not a single experience: it is a character of a huge range of different sorts of experiences that have nothing in common but that they are pleasurable, which makes judging small variations in pleasure very difficult, as in the comparison between the amount of floor-space in two very differently shaped and complicated buildings.

The very real practical difficulty, and even the practical impossibility in some cases, of judging the rank of different degrees of pleasure in different sorts of activities that is a pervasive feature of human life does not entail that any such ranking is a conceptual impossibility: it just means that value, like anything else that humans have to judge, is vague, and subject to pervasive uncertainty which can, in all but the clearest of cases, undermine very significantly the accuracy of judgments made about it. That the truth is hard to determine, however, does not make it any less true: and so, the reality that it is difficult for humans to judge which things are more pleasurable is not a meaningful criticism of a theory that holds that pleasure is of ultimate value to humans. Indeed, given that people often find it hard to decide what to do, it is more of a criticism of a theory of value that it does not entail such epistemic difficulty than that it does.

Since what is good is subject to such pervasive vagueness that so hinders the ability of humans to take accurate decisions in relation to it, how can people decide, in all but the easiest of cases, what to do? The answer lies in a theory of decisionmaking method, an outline of which is presented below.

## **Part VI – Acts, rules and decisionmaking method**

So much for a theory of the good; now for a theory of the right. As explained in Part IV above, the right is a property of particular decisions that, out of all available options, results in the most value for the decisionmaking system whose decision it is. It is often believed that such a theory of the right entails a particular sort of decisionmaking method, that is: for each decision that one has to make, and at the time that one has to make it, weigh up what the consequences are of each possible option, and choose whichever option results in what one has worked out to be the best consequences. This system of thought is often labelled “act-consequentialism”, and, not surprisingly, has many detractors. Bentham, for example, who wrote,

*“An action then may be said to be conformable to then principle of utility... when the tendency it has to augment the happiness of the community is greater than any it has to diminish it,”*<sup>76</sup>

has been branded an “act-utilitarian”<sup>77</sup>, and is subject to criticisms such as the following:

*“One objection to act-utilitarianism is that it seems to be too permissive, capable of justifying any crime, and even making it morally obligatory, if only the value of the particular consequences of the particular act is great enough. Another objection is that act-utilitarianism seems better in theory than in practice, since we hardly ever have the time and the knowledge to predict the consequences of an act, assess their value, and make comparisons with possible alternative acts,”*<sup>78</sup>

which rests on the assumption that the “weigh everything up each time” decisionmaking method is necessarily entailed by Bentham’s account of the right. However, this is a non-sequitor: it no more follows that any given standard of the right in ethics entails any particular method of decisionmaking than it follows that the standard of right answers in mathematics entails working out every mathematical problem in one’s head. That the correct solution to the mathematical equation 1+1 is 2 tells us nothing, by itself, as to how we should reach that answer: whether by mental arithmetic, or by using a calculator, or by asking a mathematically talented friend, by looking it up in a textbook on mathematics, or indeed just by remembering what we have previously been taught or calculated. Indeed, Sidgwick, adopting the same theory of the right, came to the conclusion that the correct decisionmaking method was arrived at by the development of a “moral sentiment,”<sup>79</sup> a long way from the “weigh everything up each time” method supposedly advocated by Bentham.

In reality, each practical decision is a compound decision, comprising (1) a decision about how to select decisionmaking methods in general; (2) the application of that decision in determining which decisionmaking methods are best for which sorts of decisions in general; (3) the application of that decision in determining which decisionmaking method to use for the decision in question; and (4) the application of that decisionmaking method to the decision at hand. Each of those components may, in turn, depending on the decision, be comprised of many smaller, subcomponents. It is not necessary that all of these decisions are taken simultaneously, or even that all are taken consciously, but they are all necessarily constituents of each practical decision. Superficially, it seems that this proposition is open to the following objection: since each decision is partly constituted by a decision about how to take a decision, then each decision about how to take a decision must also be constituted by a decision about how to take a decision... and so on into infinite regress. That objection does not hold, however, because only practical, applied decisions are necessarily compound decisions. Theoretical, abstract decisions (whether conscious or not), such as

---

<sup>76</sup> *Introduction to the Principles of Morals in Legislation*, ch. 1, VI.

<sup>77</sup> *Penguin Dictionary of Philosophy*; cited from <http://www.utilitarianism.com/actutil.htm>.

<sup>78</sup> *Ibid.*

<sup>79</sup> *Methods of Ethics* (Book IV, § 1).

a decision about the nature of right, and about how, in general, to select decisionmaking methods, can be simple<sup>80</sup> decisions, since no process of application of the sort inherent in the compound decisions need to be involved. To give an example, again from the realm of mathematics, the first person ever to understand that  $1+1=2$  could only have arrived at that process by a genuine, intellectual, rational appreciation of the nature of simple arithmetic. Only given an understanding of that sort can it even be meaningful to enquire which is the best method of reaching an answer to a mathematical problem. And, so, with value: to the extent to which anybody acts in consequence of conscious thought, and the second-order value of pleasure, it is necessarily true that that person, in so acting, has applied some degree of logic and rational calculation, although the logic may be flawed, and the decision heavily influenced by non-conscious (“robotic”) causes. Logic and reasoning, although often highly flawed and heavily distorted in their application, are necessary constituents of the sort of conscious decisionmaking with which humans, at least, are familiar. Therefore, the master-method, that method of decisionmaking by which all of the other methods ought to be chosen, is sound, rational enquiry. In any event, the fact of engaging in philosophical enquiry into the nature of the right necessarily assumes that, ultimately, the correct way of determining what it is right to do is by a process of reasoning. As explained below, there is no reason, however, why reasoning cannot dictate that, for some sorts of decisions at least, a process other than reasoning is the best way to reach right conclusions.

The question still remains, however, as to what rationality requires as regards decisionmaking method-selection. The answer, at least in the abstract, is not complicated: the right decisionmaking method for any given sort of decision is that which tends to produce the most right decisions most often, to the extent to which the rightness of the decisions tends to be more beneficial than the decision-costs of that method of decisionmaking are detrimental. To give an example in relation to mathematical equations again: the right decisionmaking method for simple equations might be to work them out in one’s head; for numerically complex but conceptually simple equations, to use a calculator; for conceptually complex calculations, to ask a mathematically talented friend (if available), or, if the equation is one that has already been solved, to find the solution in a book; whether to choose the book or the friend may also depend on how long that it would take to find the book, on how easy that it is to contact the friend, and so forth. The right method will also, of course, depend on the agent’s mathematical ability: a person with a talent for mental arithmetic may have a far higher threshold of complexity for calculator use than somebody who has little such ability; and, of course, somebody has to *be* the mathematically talented friend.

Likewise, more generally, the right decisionmaking method for some sorts of decisions might be to weigh the consequences; for others, to work out a rule in advance, and then defer to that on each occasion; on others, to take the advice of somebody who is more knowledgeable about the subject of the decision; on yet others, defer to some practical authority<sup>81</sup>; and on others still, to defer to some kind of intuition<sup>82</sup> (this is not necessarily an exhaustive list). Which decisionmaking method to adopt may (but will not in every case) vary, therefore, depending upon (1) the nature of the decision; (2) the circumstances in which the decision is taken<sup>83</sup>, and (3) the identity of the agent taking the

---

80 In the sense of non-compound; not necessarily in the sense of not being difficult or conceptually complex.

81 As Raz explains (for example, in *The Morality of Law* (Oxford, 1978)), deferring to an authority is the right thing to do where, by doing so, one is more likely to conform to the reasons that apply to one in any event than one would by attempting to follow the reasons directly (the normal justification thesis): in effect, Raz’s account of authority is an explanation of when one particular decisionmaking method among many, deferring to a practical authority, ought to be selected.

82 Such as where to move one’s hands to catch a ball; however, there will also be many cases where deferring to emotions, intuitions and gut-reactions are more likely to serve the values of our genes than our minds, so such a possibility ought to be taken into account when deciding whether to defer to such things, however tempting that doing so may seem.

83 This is more likely to influence considerations relating to decisionmaking costs, but it is not inconceivable that it might also influence the correctness of the outcome.

decision<sup>84</sup>. Whether it does so vary will depend on whether those factors do, in fact, tend to influence the rightness of the outcome (with due consideration to decision-costs), and therefore whether those factors are relevant to determining the best method. For these reasons, the dichotomy between “act-consequentialism” and “rule-consequentialism” is misconceived, since it erroneously presupposes that, for any given theory of the right, there is only one right decisionmaking method for all decisions.

Since practical decisions are compound, appraisals of them are necessarily also compound. A proposition that a practical decision that somebody has made is wrong is true, not only if the consequences of that particular act of deciding were worse than that of other alternatives, but also if having made that particular decision entails having made another, more abstract decision about decisionmaking method, that itself is wrong in the sense of tending to produce worse consequences (less right decisions) than an alternative that would have entailed another decision. For example, if somebody were to bet her or his life savings on a horse race, and win, there is undoubtedly a sense in which the decision was right, in that it in fact procured more beneficial consequences (assuming the possession of more money to be better than the possession of less) than the alternative of keeping the money in the building society. However, there is also a very real sense in which the decision was wrong, in that the decision to gamble one’s life savings on a particular horse entails the more abstract decision that gambling with money that it would be seriously detrimental to lose is the thing to do, which, even after the gambler’s horse luckily came in first, is still false.

Conversely, there is also a sense in which a decision taken according to the best method, but which turns out to have suboptimal consequences compared to a decision taken according to another possible method, is wrong: suppose that, instead of betting on a horse, a person put all of his life savings into a building society account. Against all the odds, the building society goes bust, and the investor loses all of his money. It is meaningful to state that, in one sense, the decision to invest in that particular building society turned out to be wrong; yet, it is also meaningful (and true), in another sense, to claim that, since the decision entailed having taken (and presumably applied correctly) another decision about how to decide how to invest one’s money *in general*; that decision, despite the bankruptcy of one particular building society, still being true, the decision was right. Even if, for some reason, market conditions had changed such that it was no longer wise to invest in building societies at all, and that one would be better off investing in government bonds, or gold bullion, there is still a sense in which the decision was right, in that the decision that investing in building societies was prudent entailed another, more abstract decision, that (for example) the best way to determine where to invest one’s money is to consult an independent financial advisor, a decision which is still right; and so on up the chain of abstractness.

There is also a sense in which a decision to criticise another decision, or to act as if it was wrong, is right, even if the decision criticised itself is not wrong, if the decision to criticise (etc.) entails another decision about how to decide when to criticise (etc.) that is itself right (and was correctly applied). However, there is also, of course, necessarily also a sense in which such a decision is wrong. The rightness of a decision as to which part, if any, of a partially wrong compound decision to criticise (etc.), in turn, is determined, not just by the rightness of the decision criticised, but by the consequences of criticising (etc.) wrong decisions of that sort, which is again a compound decision. The rightness of decisions in an interpersonal context is explained in Part VII, below.

---

84 Different people have different knowledge, skills and abilities, and may therefore tend to render more accurate decisions of certain sorts given different decisionmaking methods, even if all other things are equal.



## **Part VII – Why the good of all is the right for each**

So far, I have presented no reason to suppose that the good of any person other than the agent in question is relevant to what it is right for that agent to do: after all, an individual human mind is a single, distinct decisionmaking system, and an individual human's pleasure is its independent, second-order value for that individual. One cannot deduce that one person's pleasure, in and of itself, and all other things being equal, tends to cause another person to act in ways more likely to increase it than not, since the first person's pleasure is not part of the second's experience. Unlike for a person's own pleasure, it is not necessarily the case that any given person knows how happy that any other given person is at any point in time. Does this, then, compel the conclusion that humans all ought to behave selfishly, without any regard to the welfare of others? It turns out that it does not; but such a conclusion, obvious as it may seem to many, is not something that any theory of the right may assume; it is something that it must establish through reason, for the exact reasons why the good of all is the right for each importantly determines features about the application of such a principle that mere stipulation of such a proposition leaves underdetermined.

The reason that it turns out that the good of all is the right for each is that, in cases where what is good for the agent of decision conflicts with what is good for other conscious humans<sup>85</sup>, the best decisionmaking method is one that resolves the conflict by applying a rule that the right decisionmaking method to choose is the one that maximises the good for all those whom the decision will affect, including the agent making the decision. In other words, the good of all is the right for each because, when correctly applied as a decisionmaking method, it tends best to promote the good of each.

As explained above, each decision that an agent makes about what to do is a compound decision that entails different decisions at different levels of abstraction. Whilst the content of the right decision can vary depending on the identity of the agent taking it, at each higher level of abstraction, it is less likely that the rightness of the decision is influenced by the identity of the agent, until, at the highest level of abstraction, it is impossible that the identity of the agent can influence which decision is correct. When an agent is taking a decision in the context of many other agents, each also taking decisions on the same or related subject-matter, and in the same or approximately the same way, this logical structure entails that what is, at a sufficiently abstract level, right for each, can only be right if it is also right for each other.

For example, it might be true that, if George is feeling angry at Trevor, hitting Trevor would make George feel better. Superficially, therefore, it would appear that the right thing for George to do would be to hit Trevor. However, since each practical decision is a compound decision, for it to be true that the decision to hit Trevor is right in every sense, it must be true that all of the decisions, at each level of abstraction, that comprise it are right: the right application of a wrong principle is, after all, still a wrong decision in some relevant sense. So, for it to be the case that it is right for George to hit Trevor, it must also be the case that it is right that, in general, people ought to hit other people when they are angry at them. At this level of abstraction, the identity of the agent is not relevant, so, if it were true that it were right for George to hit Trevor whenever he was angry with him, it would also be the case that it was right for Trevor to hit George whenever he was angry with him. So, it is only the case that it is right for George to hit Trevor if it is also right for George to be hit by Trevor (or, indeed, anybody) whenever Trevor (or anybody else) is angry with him. If, in general, the pain and injuries caused by being hit when people are angry with him is worse for George than the frustration caused by not hitting people when he is angry, then it cannot be the case that it is right for George to hit Trevor when he is angry with him. If both George and Trevor experienced roughly the same pain as each other from being hit, and frustration from not hitting, respectively, then it would necessarily follow that what is right for each is that which would procure

---

85 Where there is no such conflict, of course, the question does not arise.

the optimal consequences for all, since what is optimal for each would also be optimal for each other.

Suppose, conversely, that, whilst Trevor was none too fond of being hit, George was a masochist, and quite enjoyed pain. Would it now be true that it is right for George to hit Trevor whenever he is angry with him, since that would entail it being right for Trevor to hit George sometimes, which is the optimal outcome for George, since he enjoys being hit? That would only be right, in turn, if the yet more abstract decision that it entailed, that each agent ought act so as to maximise the things that he or she finds pleasurable, without regard to what others may find differentially pleasurable, is right. That decision, being again sufficiently abstract so as to make the identity of the agent irrelevant, would only be right if the consequences to each (including the agent taking the decision) of all acting on it were more optimal than the consequences to each of all (including the agent) acting on a contrary decision, that is that what each agent ought to do is maximise the good for all. Since it is highly probable that, in each case, greater benefit accrues to each agent by each other agent's acting to promote maximal good than detriment accrues to each agent by so acting, the right thing for each agent to do is to promote maximal good.

In other words, where different states of affairs are differentially valuable to different rational, intelligent, forward-thinking decisionmakers, the right decision at almost the top level of abstraction for each decisionmaker is to adopt whatever decisionmaking method most tends to maximise the overall value of states of affairs in general amongst all such decisionmakers, since the rightness of the decision at that level of abstractness entails that it is also right for each other decisionmaker, and each decisionmaker's good is most effectively promoted when each other decisionmaker takes decisions on that basis, even taking into account any detriment entailed to any given decisionmaker in maximising the good of others, at the expense of her or his own. In short: the good of all is the right for each because the good of all is the good for each.

I suspect that some will, to this principle, raise the following objection: if the good of any given agent is best promoted by, for example, maximising the good of some subset of the world's population of humans, rather than the world's population as a whole, and where that is also true of the other members of that subset, then is the right thing to do not, after all, to act to promote the welfare of each other, but only of those in the subset? The answer, as with all such objections<sup>86</sup>, is this: for it to be right that any given subset of people maximise each other's, rather than everybody's, welfare, it would have to be right that such was true of each such possible subset. It would only be true if the more abstract principle, applicable to all, whether in any given subset or not, that people should, in general, promote the good of some, but not all others (at least in some circumstances), tended to promote the good of each more than the contrary principle, that people should, in general, pursue the good of all. Necessarily, of course, this could not be the case unless attempting to promote the good of only subsets *did*, in fact, best promote the good of all. Thus, such a contrary principle is either untrue, or not contrary at all; and so, with each case where a person might attempt to calculate how her or his own good might better be pursued by not pursuing the general good: doing so would only be right if the abstract decision, applying to everybody that it entailed, that making such calculation, in general, is the right thing to do, entailed the best consequences for each agent. Since it will not be the best for each unless it is also the best for all<sup>87</sup>, it cannot be the case that any such decision is, in each part that comprises its compound whole, right.

This principle also explains why it is the case that the right consisting of that which maximises the good of all does not require, as some have suggested, that each agent be as altruistic to others as he or she possibly can: such a decision would also, in a relevant sense, be wrong, since it could only be

---

86 Of which, no doubt, a very large number could be conceived: I use this as but one example of the genre.

87 Since, at the requisite degree of abstractness, each means each of all.

right if it was the case that the maximum possible altruism benefited each in the receipt more than it caused detriment to each in the performance: since it is likely to be true that, overall, people will benefit more from not spending every waking moment doing nothing other than helping others than they would benefit from everybody else doing the same<sup>88</sup>, it is also likely to be true that the right thing to do on many occasions is that which is in the agent's own interests, even if another decision would best promote the welfare of some other or others.

So, the rational thing for each to do is to maximise the good for all. This would be so even if it was not the case that humans often obtain pleasure directly from others' pleasure, or feel displeasure in direct consequence of other's pain. All other things being equal, many people (although not necessarily all people) tend to prefer that others be more rather than less happy<sup>89</sup>, and such a preference can (but will not necessarily) make it right for such people to be more altruistic towards other people than is entailed by the greatest happiness principle operating on people who do not have such a preference alone. Since the preference for others' happiness is often stronger in relation to some, rather than other people, such a phenomenon can also make it right for the people who are subject to it, in some circumstances at least, to treat, for example, their friends and family, in a way more favourable than other people. This does not, however, entail that such favouritism is always permissible where such a preference exists: it is only right to act with favouritism where, in general, the benefits of so acting to the agent outweigh the detriments to others, less the benefits for the particular others thereby benefited. So, for example, it may be right for a person to give birthday presents only to close friends and family, but to decide between applicants for jobs solely on the basis of merit.

Since each person is likely to have significantly less information about each other person's pleasure, and the things that tend to give rise to it, than her or his own, it also follows that it different sorts of decisionmaking method are right for decisions that have a significant impact on others' pleasure than those that only impact on the decisionmaker's pleasure. It is likely that, where others are significantly concerned, right decisions are more likely to be arrived at by applying more cautious, conservative decisionmaking methods (such as rule-following, or deference to authorities) than where only the decisionmaker her or himself will benefit or suffer detriment by the decision, where less cautious, weigh-everything-up-each-time sorts of methods are more likely to produce correct decisions. This does not mean that, in *all* cases where others are involved rule-following or authority-deferring decisionmaking is to be preferred, and in *all* cases where others are not involved, calculation of benefits on each occasion will be better: merely that, all other things being equal, it is more likely that more cautious methods of decisionmaking will be the optimal methods in cases where others' pleasure must be taken into account than where that is not the case.

Thus, the "greatest happiness principle" long advocated by the classical utilitarians such as Bentham and Mill, far from ignoring the "separateness of persons" as Rawls<sup>90</sup> alleged, inevitably follows from due recognition of it. Common criticisms of it: that it is too demanding, that it pays insufficient regard to reasons for action specific to each agent, that there is no particular reason for advancing everyone's, rather than one's own (or specific subsets') happiness, that it does not account for the distinction in decisionmaking method between cases involving the only the agent's own welfare, and those others' welfare, and that, as a decisionmaking method, it tends to induce suboptimal outcomes, are therefore more illusory than real. However, one important caveat appears, which limits to some extent the scope of the principle: it is explained below.

---

88 After all, what opportunity would anybody have to receive the benefit of such altruism if everybody was permanently engaged in providing it for others?

89 It is not necessary to dwell on the reason for that here, although one might readily speculate as to such a tendency's evolutionary origins.

90 Supra.

### **Part VIII – Justice, law and punishment: dealing with the irrational**

It hardly needs mentioning that not all humans are particularly rational, intelligent and forward-thinking. People get decisions wrong very often; sometimes, such errors only affect the decisionmaker (such as in the case of the imprudent investor)<sup>91</sup>, whereas on other occasions, wrong decisions affect others, too (as in the angry George example). The question then arises about what it is right for each agent to do about the wrong of others. As always, the correct method is that which is correct at every level of abstraction, including levels of abstraction where the identity of the agent is not relevant; and this entails that the right for each is the good of all who take the right decision at that abstract level. Thus, where the good of all such people is maximised by doing nothing in response to a wrong decision, nothing ought to be done; where it is maximised by educating people so as to make them tend to be more rational, this ought to be done; and where it is maximised by punishing all those who do a wrong of any given sort, this ought to be done. In all of the cases (except where the right thing to do is nothing), adopting any particular method (punishment, education, or anything else) is only right if the costs of administering it are outweighed by the benefits of its success. This, in turn, entails that, the more effective the measure, in general, the more costly that it may be. In turn, the decisions about when inaction, education, punishment, or anything else, is the right response to any given sort of wrongdoing, ought to be taken on whatever basis, both in the abstract and in the application, is the most reliable way of reaching such decisions correctly.

There is not space in a paper of this sort to write in detail about what such principles entail: such an answer would in any case depend heavily upon empirical, contingent considerations, evidence about and tools for analysing which are not readily at the disposal of the philosopher. However, some general comments about the sorts of things that such a principle is likely to entail may assist in elucidating the foundation of the principle itself.

In relation to punishment, the principle is that, in many cases, the form of irrationality or unreasonableness in those who do wrong is failing to appreciate what their decision entails at a higher level of abstractness; if it is possible, therefore, for people between themselves to organise a system whereby, whenever a person does a wrong of certain sorts, the likelihood being that the wrong was done for personal gain (either calculatedly, as in the case of most theft, or unthinkingly, as in the case of angry Georges), loss significantly greater than that gain be inflicted upon that person so that, those who know of the existence of this system will, even if they fail to take into account the abstract component of their decisions, nonetheless tend to conclude more often than they otherwise would that the wrongful action is not one that they will seek to perform. This is justified where both the detriments of imposing the punishment are outweighed by the benefits in tending to prevent the detriments of the wrongdoing to which the punishment is a response. As with all such calculations, this must include the likelihood and degree of probable adversity of the consequences of erroneous applications of the correct method.

Whether to punish at all, the nature of the punishment, and the system of its administration, will therefore vary depending on the sort of wrongs sought to be prevented. Generally, a more costly system of administration (and, where relevant, detection), and punishments more costly to the individual punished, are on this basis justified for wrongs that tend to cause more serious detriments; and a less costly system, and less costly punishments, for less serious wrongs. So, for important matters, the principle would be likely to justify the establishment and maintenance of a criminal justice system<sup>92</sup>, with expensive mechanisms of enforcement and, at least potentially, quite

---

91 Of course, such wrong decisions are perfectly capable of affecting others; indeed, any decision that anybody makes eventually effects most things; but, in this context, a decision affects others where the nature of the decision itself is likely to make a significant difference to the extent to which states of affairs are valuable in relation to other people, in a consistent manner.

92 Of course, law can serve purposes other than this: nonetheless, its ability to serve such a purpose is an important

severe punishments. For less serious matters, such as discourtesy, a far less formal system of punishment, entailing far less unpleasant consequences for the wrongdoer, by way, for example, of social disapprobation, is likely to be justified. In either case, the costs of inflicting such punishment are only justified where it is likely to be effective, and they can only be effective where knowledge of the possibility of punishment can have an effect on a person's decisionmaking. Thus, for example, in the case of the mentally infirm, or those who, like the prudent but unfortunate investor, have made a wrong decision that was nonetheless constituted by the correct application of the correct method (in other words, a reasonable decision; or, in language more familiar to lawyers, a decision without fault), responses other than punishment (such as secure hospitalisation, or inaction) may be more correct.

This also explains the first part of the principle of justice, expounded summarily in Part IV above: that no person shall benefit from her or his own wrong. Justice is the principle that, in the abstract, deals with antecedent instances of wrongdoing, and as noted above, cannot be fully elucidated here, for this is but a brief outline of what this theory of the right entails about those who do wrong. Nonetheless, it follows that justice requires more, rather than less, effective methods of preventing wrongdoing, even in many cases where this is to the significant detriment of the wrongdoer; for, since, save for the possibility of errors, acting rationally will almost inevitably exclude the possibility of any rationally-imposed punishment, the potential detriment to the punished need feature only to a small extent in the calculus of the rational in deciding whether to impose such a punishment. Indeed, were it possible to eliminate entirely the possibility of error, it would inevitably follow that the unpleasantness of the punishment imposed on the irrational is irrelevant to the right thing to do, except in so far as is relevant to the effectiveness of the punishment, for the rational could be sure to avoid such consequences, since doing the right thing would inevitably entail such avoidance.

This leads to an interesting conclusion, and one of the few in this paper that is contrary in nature to central tenets of classical utilitarianism; that is: whilst it is true that, in a world of wholly rational people, the right thing to do will necessarily tend to benefit all, the principle of the right itself entails no more than doing that which maximises the good of people *only in so far as they themselves do the right thing*. Thus, when Bentham wrote,

*"...all punishment is mischief: all punishment in itself is evil. Upon the principle of utility, if it ought at all to be admitted, it ought only to be admitted in as far as it promises to exclude some greater evil,"*<sup>93</sup>

he made an error; for the only reason for taking into account the pleasure of other people is that failing to do so would entail that it would also be right for others to fail to do so, which would be detrimental to each agent. In a case where a person acts wrongly, however, that the right decision entails not taking account of that person's pleasure does not entail that the agent taking that decision is also likely to suffer similar detriment as a result of others' right decisions, since the agent, in doing the right thing, necessarily avoids the prospect of rightly-imposed punishment.

This does not, however, entail that the smallest wrongdoing requires the greatest punishment, since, in reality, there is always a considerable chance of error at all levels of the procedure: in calculating which things are wrong, in determining whether a wrong thing has been done, in and identifying who it is that has done wrong, mistakes are always possible, and, given the sample size of a sufficiently large society, overwhelmingly probable to occur repeatedly and irrepressibly. The chance of error, of course, entails a chance, and not an inconsiderable chance, that rational actors may nonetheless find themselves subject to punishment imposed by other rational actors. Furthermore, there are costs associated with punishment, both in administration, and, in the cases of

---

reason to have law, despite the costs of establishing its necessary institutions.

<sup>93</sup> *An Introduction to the Principles of Morals and Legislation*, ch. XIII, para. II.

serious punishment, especially imprisonment, the loss of productive labour from the economy; generally, the more severe the punishment, the greater the cost (although the converse is true in the case of fines, of course). In relation to less formal, non-legal, punishments, the cost in terms of the displeasure to those administering the punishment, for example, of persistent antipathy, is often significantly higher than the cost of less effectively checking, say, discourteous behaviour, and so, again, the most severe possible punishment is not always justified.

Furthermore, such a proposition does not entail non-optimality: for the most optimal outcome on either this principle, *or* the unmodified greatest happiness principle is identical, and can only come about by all agents acting rationally (or by highly improbable chance). In so far as this theory requires the most effective possible way (taking into account the costs) of compelling all to act as closely to the optimal model as possible, acting according to this theory of decisionmaking method is more, rather than less, likely, I should suggest, to produce an outcome closer to the optimal than a decisionmaking method based on the unmodified greatest happiness principle.

Of course, cliché tritely but correctly prescribes that prevention is better than cure: and so, despite the potential costliness of doing so, it is more efficient to attempt to cause people to think more rationally in the first place than to have to punish them for being irrational later, at least to the extent to which such an attempt is effective. Such a proposition is, I suspect, fairly uncontroversial, and entails things like a good education system for children, and that parents ought to teach their offspring how to behave, and so forth. Indeed, I should go so far as to suggest that primary school children ought to be taught philosophy in order to increase the likelihood that they will, as adults, tend to think more rationally, and make better choices, although, of course, the correctness of that opinion depends on empirical contingencies beyond the scope of this theory itself.

As stated above, how rational people doing the right thing ought to deal with irrational people who do the wrong thing is a topic far more vast in scope than is capable of being covered, except at the most abstract and general level, in several pages in a paper on a theory of value. Even accepting this theory, there are a multiplicity of debates to be had about what sorts of decisionmaking methods do, in fact, tend to produce the most reliable decisions; what sorts of costs do, in fact, outweigh the benefits of controlling particular sorts of wrongdoing, what sort of wrongdoing ought be controlled at all, and which particular things are wrong in the first place. One of the central reasons for the persistence of such debates is the pervasive epistemic difficulty that accompanies all forms of evaluation, and is multiplied as the complexity of the judgments to be made multiply, especially when the differing interests of thousands, or even millions, of people are involved. As will be obvious from the theory of compound decisions, even the most straightforward of evaluative decisions are actually extremely complex in some respects; it is little wonder, then, that many, when faced with this complexity and uncertainty, turn to intuitions and other non-rational systems of thought when faced with making such choices.

## Conclusion

The complexity of the right and the vagueness of the good has probably been what has lead many astray in their search for understanding of both. Compounding that, of course, is that it is only in the last century or so that a full understanding of life, and its evolutionary origins, has made possible the analysis of value presented herein, the logical structure of which necessarily depends on it. At the time that Hume wrote that value cannot be illuminated by reason, the nature of life, and, therefore inevitably, the nature of value, was not fully understood, in consequence of a lack of empirical investigation.

It is hardly surprising, then, that philosophers before Darwin's time were unable to make significant progress in expounding theories of value which explained its nature in anything approaching a conclusive and independent manner<sup>94</sup>, or that did anything more than investigate that which is entailed by unreasoned but common assumptions, such as that those about the relevance of intuitions. What is more surprising, however, is that, although over a century has passed since the publication of *The Origin of the Species*, moral philosophers have largely continued as they were before, without reference to the theory that explains as evidenced and contingent the nature of life, for which they all inevitably agree<sup>95</sup> that everything good is valuable. Perhaps it is the lamentable and arbitrary dichotomy between the arts and the sciences that has lead to this lack of congruity between the two necessarily closely connected fields; or perhaps it is the fear of a naïve, simple, and, ultimately, false theory of value<sup>96</sup> emerging from evolutionary theory that has kept moral philosophy in the intuition-based (or sceptical) dark ages<sup>97</sup>.

What is perhaps also surprising, although more pleasantly, is that philosophers such as Bentham and Mill, writing, in Bentham's case, nearly a century before Darwin's revolutionary discovery, were able to fashion theories extremely close to that which an empirical investigation, in the light of modern scientific understanding, can show to be true. Perhaps it was luck; or perhaps it was a form of critical self-awareness of their own motivations to action that inspired the early classical utilitarians to understand the nature of value and the right, as applied to humans, so well without having before them the evidence necessary to understand its logical structure and relation to life. Perhaps this theory, too, will need modification in the light of future scientific discoveries about the nature of and relationship between pleasure, desire and motivation; but, even with such modification, the theory that can most accurately explain both the nature of value, and that which is valuable, will always remain that sort of theory that is uniquely able to explain the nature of contingent facts in the world, such as value: an empirical theory.

---

94 Independent in the sense of explaining value in non-evaluative terms.

95 As explained in Part II above

96 Such as that which G. E. Moore falsely accused Herbert Spencer of propounding, in *Principia Ethica*.

97 Such a fear was almost certainly behind Dawkins' warning (supra) in the opening chapter of *The Selfish Gene*.