Ronald de Sousa

sousa@chass.utoronto.ca

University of Toronto

October 2001

ABSTRACT

Many psychological theories of emotion stress their motivational power. But motivational power belongs primarily to desires. What then differentiates emotions? Since for any possible act, one can only choose to do it or refrain, the assimilation of emotion to desire is a projection of a multidimensional reality onto a two dimensional plane--a reduction of colour to black-and-white. To explore the richer, full-colour domain of emotion, which I refer to as the "axiological", I shall turn first to the role of emotions in learning. Learning, whether of skills or of knowledge, typically extends the range of our pleasures. This can be plausibly regarded as a biological adaptation. Emotions are essential to learning intellectual and bodily skills: some, like curiosity, motivate inquiry, others, like pride, reward its outcome. Moral education also involves our emotions, but the range of relevant emotions is broader -- including, perhaps, intrinsically moral emotions such as compassion, indignation, empathy, as well as others, such as jealousy, spite, that we are inclined to classify as intrinsically nasty. Insofar as most of those emotions are learned in early formative stages of moral development their effect is probably inherently conservative. Here, though, art and literature may succeed in educating our emotions into greater subtlety and complexity. Emotions as they are revealed by great literature do not function merely as desires, prompting calculation of means to their satisfaction. They reveal a multi-coloured reality which contributes to axiological knowledge and ethical growth. In this paper, I shall try to improve my understanding of how that can be so.

Emotions in Black and White

Insofar as emotions are required only to *motivate*, we can think of them as just pushing and pulling: swayed by the totality of your emotions at any particular time, either you act or you don't. Complications affect the way we *experience the world*, but in the end they have to be funneled into a sequence of single decisions, where each requires acting or not acting, the black-and-white of yes or no. If we focus on the experience of emotions, on the other hand, they are so diverse as to constitute no single *kind* of thing at all. Each carries a wealth of specific meanings enriched by an immensely large class of contrasts: that is the polychrome vision. The monochrome vision comes in the context of action and choice; the full-colour vision is of emotions as modes of multi-dimensional *axiological perception*.

On the full-colour view of emotions, it could be argued that there are no practical limits to the number of distinct emotions that can be experienced, any more than there are limits to the number of thoughts one can have. This view seems plausible in the light of the emotions aroused by aesthetic experience. In watching dance or listening to music, for example, it seems obvious that emotions are involved in some way, and yet this observation is threatened with two unequal and opposite forms of triviality. On the one hand, one may be tempted to think that the emotions expressed by music and dance are "the great emotions" -- those we can list on demand: anger, fear, love, awe, jealousy, sadness, desire. But then *why go to all this trouble?* If works of art exist merely to evoke those emotions, and if there are no significant differences between any two instances of "fear", or "anger", and so forth, it hardly seems likely that works of art in all their diversity should be sustaining our interest for their representation of *emotion*, rather than for some other reasons. On this view, it's difficult to see why most art isn't superfluous. (Most mediocre art is indeed superfluous in just this sense: once you've seen one scary alien-invasion movie, you've seen them all.) The alternative view is that each different moment in music or

dance (or any other art that is thought to evoke emotion) is actually expressing an emotion *sui generis*, or rather not just sui *generis* but, as it were, *sui ipsius:* in its own unrepeatable individuality. But then the role of art is trivialized again: for if every piece of art necessarily expresses its correlative emotion, no more and no less, then that seems to remove the possibility that some forms of art might be better or worse at expressing emotion, or that the emotions evoked by some work of art might be more worth-while, more interesting, more deeply felt, more authentic than others. For all those evaluative judgments seem to presuppose that there is something beyond the expression of emotion in terms of which the latter can be judged.

This last objection can be overcome, if we give up the assumption that each work of art or literature is seen as conveying a ready-made emotional "message". If instead art is thought of as creating and embodying a particular emotion of its own, then there can be numberless emotions, and every work of art is more or less interesting in accordance with the quality of the unique emotion it conveys. There are then indeed literally *innumerable* emotions.

But now problems of a different sort arise.

If the emotions are indeed innumerable, how are we to understand their influence on our choices? The question is premised on the assumption, which many of us have learned from Aristotle, that moral education is in large part a matter of learning to feel the right thing in the right way at the right time. (Aristotle, NE 1104b13) In the course of moral education, we somehow learn to feel certain emotions; in turn, in our later life, those emotions (a) allow us to *see and understand* things correctly and (b) *move* us to act rightly in response to that understanding. But how are we to think of them as performing that mediating function?

One promising answer focuses on the importance of literature in our moral education. This has been argued forcefully by Martha Nussbaun, who has stressed the capacity of literature to present us with "fully imagined" emotions and particular characters, as contrasted with the bloodless abstractions customarily paraded in philosophical examples (Nussbaum 1990, passim). This idea, however, raises two problems. First, it however detailed and vividly imagined a fiction

may be, it can't ever be literally true that fictions present us with particular individuals. Concrete particulars have an infinity of relational properties. Fictions inherit the intrinsic generality of the words in which they are conveyed, and their determinate properties are necessarily finite. They can be indefinitely *specific*, but never *particular*. Second, if we concentrate on the unique features of each situation, on the specific qualities of each individual, and on the singularity of each emotion, it is difficult to see how they could provide any guidance at all. Learning requires comparisons, which in turn require similarities and differences, classified and conceptualized in terms that necessarily return us from the particular to the general. Learning, in short, requires repeatable patterns. Taken literally, then, the suggestion that each particular situation in life (as well as each episode in a work of art) gives rise to an unrepeatably unique emotion is self-defeating. (Particular) percepts without (general) concepts cannot guide our choices.

This problem arises not only for individual learning, but also for the evolution of our emotional capacities by natural selection. While some social scientists and philosophers are keen to minimize the contribution we can expect biology to make to our self-understanding, few would deny that natural selection has had a hand in determining our potential for emotions. Certain emotions, sometimes labeled "basic", have obvious relevance to vital functions, plausibly served in some species by instincts, such as those sometimes referred to as the "Four Fs" These obviously couldn't have been selected if they had not given rise to repeatable patterns with repeatable causal properties.

Clearly, then, the mediating role of emotions must depend on some common features. In the past I have suggested that beyond the half-dozen or so patterns commonly listed by emotions theorists, such as those that Ekman has found to be more or less universally recognizable in their mode of expressions, the common features of more complex emotions are embedded in *paradigm scenarios*, which determine not only different types of emotion, but their characteristic felt quality as well as their conditions of appropriateness. The defining roots of each distinguishable emotion lie in a specific dramatic situation or episode type, associated with a

characteristic *feel* (in the broad sense, not in some specific sense limited to bodily *feelings*). The drama is played out by agents in certain *roles*, and the hypothesis that formed my starting point was that these paradigm scenarios, which can vary from one person to another, actually define the *meaning* for us of the emotions of which they constitute the prototypes. When such an emotion is experienced in later life, its appropriateness to the situation is determined in part by the relation between the present situation and the original paradigm, which determines the *formal object* of that emotion, where the formal object determines the standard of appropriateness that must be met for an emotion to be that kind of emotion. As a story can be told in more or less detail, so emotions can be classed together as variously fine-grained determinables and determinates. On this view, it's not a problem that very different experiences can all count as fear, or as anger. While the plot of *Antigone* might be summarized in a line (*Princess quarrels with uncle, dies with fiancé*) the impact of the play is not explained by that summary. Similarly, the determinate emotions we experience are not adequately described by the standard words we use to refer to determinable emotions.

This suggests a way of reducing the variety of experienced emotions, insofar as most of our emotions are importantly *repetitive*, and classified in part in terms of a finite repertoire of *roles*. Yet it does not seem sufficient to answer the problem I have posed: how is it that emotions mediate our moral learning?

In order to provide that answer, I suggest we start by returning to the monochrome perspective: instead of considering emotions in their variety, look at them as projected into the simplifying single dimension of aversion/attraction. As when a full-colour picture is reduced to a black and white drawing, much wealth of detail may be lost, but some clarification may result. The simplest version of the monochrome perspective is the much reviled hypothesis of psychological hedonism, which I propose to revive in a version I'll call *immediate hedonism*. Later, we shall see just what additional requirements call for an expansion into the full-colour perspective of axiology.

Immediate Hedonism

Psychological hedonism is the doctrine that we act, as a matter of fact, only in pursuit of pleasure (and avoidance of pain). This must not, of course, be confused with *ethical hedonism*, which is the view that pleasure (and avoidance of pain) ought to be pursued. The received wisdom about these two doctrines is that they are incompatible. If we can't help following our inclinations to pleasure, it can hardly serve any purpose to tell us that we ought to do just that. Yet there are ambiguities in each doctrine that might be resolved in such a way as to allow them to be held together.

The first ambiguity is this. Psychological hedonism says that our ultimate motivation is always pleasure: but whose pleasure? It is assumed that we mean: my own, for every agent. Ethical hedonism might view the value of pleasure more broadly, and merge into the utilitarian doctrine that one ought to aim to maximize pleasure for the greatest number. This may indeed be incompatible with the exclusive pursuit of my own pleasure -- and so it may propose an ideal that each agent is actually ill-equipped to aspire to. But being ill-equipped to aspire to it is not the same as trivially conforming to it. The second ambiguity concerns psychological hedonism alone. Is it a question of maximizing my pleasure over all, or of invariably pursuing whatever pleasure is most accessible at any particular moment? The latter version is something like the policy of "pursuing the present pleasure" attributed by Aristotle to the "akolastos" or "self-indulgent". It leaves room for a version of ethical hedonism that doesn't even take others into account, but merely advocates taking a broader view of my own.

Among the objections that have been leveled against psychological hedonism is the one that invokes the thought experiment of the *experience machine*. (Nozick 1974, 42-45)

Subjects are asked to imagine that they are offered the choice of continuing life as it is, with all its uncertainties and risks of pain, or to agree to being plugged into a machine that is guaranteed to give them, for the rest of their lives, only pleasure -- or rather, if that sounds altogether too dull, just the right mix of moments of anticipation, frustration, and final success,

just the perfect mix of moments of hope, anticipation, and ultimate fulfillment. Most people, it turns out, reject the machine. If we discount the boring hypothesis that they simply can't actually imagine believing the promise, this poses a problem for psychological hedonism. But as Sober and Wilson have recently pointed out, this argument is without force against a version of psychological hedonism which interprets it as geared entirely to the present pleasure, rather than to the results of calculations aimed at maximizing future or overall pleasure (Sober and Wilson 1998, 275-291). This is the version of psychological hedonism I shall refer to as *Immediate Hedonism*.

Immediate hedonism accords the decisive causal power to immediate pleasure, that is, in particular, to the actual experience of making a given choice, rather than to the consequences of the choice for the future. It can easily take account of the problem of the experience machine. For it need only postulate that the act of choosing real life is in itself more pleasant than the act of choosing the experience machine, however much more pleasant the consequences of the latter choice might seem.

One advantage of this hypothesis is that it allows for the possibility of a choice the consequence of which would be less pleasant than the consequences of the choice not made. For the following proposition (1) implies no contradiction:

1. *A* is preferred to *B*, but *choosing B* is preferred to *choosing A*.

So immediate hedonism escapes the most convincing argument against the usual versions of psychological hedonism.

That isn't to say that the doctrine is unobjectionable. It can still be charged with a kind of triviality. But before we yield to the temptation to dismiss it on those grounds, note that it makes room for the view that while (1) is consistent, it is also *irrational*.

The Philebus Principle

The reason (1) is irrational is that it violates a plausible principle of rationality that might be called the *Philebus principle*.

In the course of a thorough investigation of pleasure, Plato's *Philebus* explores the possibility of treating pleasure, in effect, as a *species of perception*, and not merely a species of *sensation.* The significance of this distinction is that perception, unlike sensation, purports to convey information about something other than itself. In particular, pleasures of anticipation focused on some future pleasure may provide information as to the extent of that future pleasure. This may not be sufficient to regard them as literally true or false, as Plato claims, but it does seem to support the following [PP] as a principle of rationality:

[PP] *A **pleasure of anticipation** should be proportional in intensity to the **anticipated pleasure** which it depicts.*

From a biological point of view, this is a reasonable principle. On the assumption that Immediate Hedonism is true, the intensity of the present anticipatory pleasure will determine the tenacity of my present motivation to pursue the pleasure anticipated. But if my present pleasure is uncorrelated in intensity to the quantity of future pleasure which it motivates me to pursue, then my present motivation will bear no settled relation to me future satisfaction. Our lives would be perpetually like those of a person who, in planning a picnic for a specific place and time, takes meticulous account of a weather forecast for a possibly different, randomly chosen place and time.

Emotions and Learning

We can now discern some of the way that emotions, as projected onto the "monochrome" dimension of pleasure and plain, could serve to motivate our progress in various domains. They do so, in fact, in at least three different ways.

Most obviously, Aristotle's observation that every human being by nature desires to know (Met 980a21), is a correct empirical observation which applies from the earliest days of childhood. Pleasure is involved in the exercise of all competencies, and in learning to increase the power of our competencies. All children manifest this sort of pleasure, at least until schools

curb their curiosity and turn it to revulsion. The pleasure of learning, from an evolutionary point of view, is no more surprising than the pleasure of sex.

Secondly, this pattern also gives us one way of conceptualizing akrasia: akrasia can be viewed as the choice of A rather than B at the very moment when one grants the preferability of B over A. This, as I've noted, must be both possible and irrational, and it can be thought of as exemplifying the very same pattern: at the moment of choice, I envisage the prospect of B as preferable -- that is, on our present restricted, black-and-white, two-dimensional schema, affording a greater quantity of pleasure -- in comparison with A. But because of what Plato might call a false pleasure of anticipation, I experience more pleasure in the actual act of choosing A than in the act of choosing B.

Now if what I've said so far is right, this represents not only a possible pattern but also an irrational one. But where there's irrationality, there's hope for rationality. The bright side of akrasia is that it bears witness to the possibility of moral and emotional progress. Progress implies conflict: it presupposes that in matters of moral behaviour as well as experienced emotion, it is possible both to make mistakes and to regret them. Insofar as such mistakes can be corrected in the future, they herald improvement. (Aristotle rightly notes that his 'akolastos', who "pursues the present pleasure" and just does the wrong thing without experiencing any inner conflict, has no regrets and is therefore "incurable" (EN1150b29). He is a sort of happy moral idiot, beyond the reach of reform.)

So our restricted, black-and-white picture of Immediate Hedonism already goes some way towards explaining how some learning is facilitated by emotions. There is also a more direct, purely physiological route, which is neutral between theories of motivation because its effect is purely causal rather than motivational. The sheer physiological stimulation of the limbic system plays a part in the fixing of a contemporaneous event in memory (Rosenfield 1992, 165), as every American knows who remembers precisely where she or he was on hearing the news of John Kennedy's death. (Assassinations having since become routine, they are no longer memorable.)

The stripped-down black-and-white picture of motivation as pleasure also allows for a range of emotions to influence behaviour and learning through instrumental calculation. The emotions that are capable of working in this way include emulation, admiration, pride, and no doubt also a number of emotions which we might otherwise be reluctant to avow, and may well wonder whether we might be better off without: greed, envy, vanity, jealousy, humiliation, and fear are all capable of spurring us on to efforts likely to lead us to broaden the scope of our competencies. In this way, they are likely to extend our learning beyond the range of those things that we learn just because the process of learning them is sheer pleasure. There are plenty of those, alas, beginning with the toil of learning to read, which every child and every parent must have occasionally thought of as a form of torture, but which the parent most concerned for the welfare of her child is least moved to spare the child. But these emotions are incidental to the promotion of such worthy ends; there doesn't seem to be anything about them that intrinsically destines them to serve the end of cognitive, emotional or moral progress.

Moral Emotions

There are, however, some emotions that are traditionally thought to do just that, emotions that one might be tempted to call intrinsically moral. Chief among these are benevolence and compassion, but others, such as guilt, shame and indignation might also figure on the list. This last, in particular, was thought to be the crucial moral emotion by François Itard, the XVIIIth century doctor who reported on the wild child he had found in the woods of Aveyron. Itard set up an experiment designed to test whether the wild child had acquired the sense of justice. This experiment consisted in "punishing" him for doing something correctly, and Itard was gratified to find that in contrast to his usual acceptance of punishment, the child reacted with fury and indignation (Itard 1964, section 54).

The most interesting aspect of this experiment is that since the wild child remained unable to speak, the interpretation put on his behaviour by Itard presupposes that he could clearly distinguish the expression of indignation or fury from that of shame, sadness, or frustration. It

presupposes, in other words, that those emotions and their expressions are natural and pre-linguistic. And in this Itard was probably correct, according to the subsequent observations of Darwin and more recently of Paul Ekman (Darwin 1998) which seem to establish that for about half a dozen emotions, including anger and fear, the relevant facial expressions are clearly identifiable across cultures. But it's hard to know how the case of a child's attitude to punishment would generalize to other aspects of justice. The absence of language in this case makes it particularly difficult to determine whether the child's outrage would have generalized to offenses against others than himself, or indeed might have been mobilized against himself in cases where he was the one that stood to lose by a just distribution. It seems likely, however, that where language is available such "natural" sentiments could be used to build a more universal sense of justice, one which meets the test of altruism.

Altruism is a characteristic that originally applies to motivation, and inconclusive arguments have raged as to whether people are ever actually altruistic or whether they "ultimately" help others in order to secure the satisfaction they happen to derive from doing so. The inconclusiveness stems from the fact that even if I genuinely desire the good of another, nothing precludes my experiencing pleasure at the attainment of that good by the object of my concern. The doctrine of immediate hedonism which I have been using to frame my argument so far is clearly committed to the view that the pleasure taken by the agent in the act is the direct motivating cause of benevolent actions, but it is entirely compatible with the complete genuineness of my concern for the other. For that reason, the debate doesn't seem very interesting.

In recent years, altruism has also been given a precise definition within biology, as a type of behaviour that results in a net loss to the agent's own fitness coupled with a gain in fitness for others. So understood, it presents a paradox for Darwinism and so raises far more substantial questions. For the challenge that Darwinism sets for itself (briefly, that any heritable tendency to lower one's fitness relative to the fitness of another must necessarily breed itself out of

existence) requires that we uncover mechanisms capable of bringing it about that organisms behave altruistically. Hence the panoply of kin selection, reciprocal altruism, and models of group selection which have been the object of fierce disputes in the literature in the past couple of decades. In terms of this concept, we can ask whether evolution might have honed in us certain emotional dispositions that were actually meant to function to promote not just sociality but altruism (Sober and Wilson 1998).

The evolutionary point of view, as Sober and Wilson have observed, "requires us to regard the human mind as a proximate mechanism for causing organisms to produce adaptive behaviors." (Sober and Wilson, 1998, 200) In particular, intrinsic or ultimate desires, whether or not they turn out to be invariably self-regarding, seem to be psychologically remote from the concrete means of their satisfaction, yet they are the proximate means by which selection secures its "end". In that light, the emotions on my list of plausible "moral emotions" might have been, in part, selected for that end.

As Martha Nussbaum (1996) has argued, compassion is among a small number of named emotions that have a special value as specifically moral emotions: emotions likely to promote moral behaviour in themselves. The sociobiological perspective to which I have been partial so far sees no methodological vice in looking to the net overall effect of a certain type of behaviour for a clue as to its function. But the principal lesson that we can learn from Nussbaum is that we need to see that the way compassion functions in a far broader context. We need to see it in full colour, that is, in its real multi-dimensional context. To sketch why this is to will be the main task of what follows.

Going Full-Colour

Nussbaum surmises that "the pain of another will be an object of my concern only if I acknowledge some sort of community between myself and the other, understanding what it might be for me to face such pain." (Nussbaum 1996, 35) And again, "the judgment of similar possibilities is part of a construct that bridges the gap between prudential concern and altruism."

(36). I wonder, however, whether this may not have it backwards. I wonder, that is whether a certain natural capacity for experiencing another's pain, in a way that actually moves us to take it into account, might not be part of the cause of our capacity for what seems a far more sophisticated mental act of "acknowledging community." I don't see why we shouldn't take at face value the reports that point to a capacity in some of our mammalian cousins, such as whales, porpoises or elephants, for empathizing with members of their own or even other mammalian species (Masson 1995, 79-82). But to accept this observation at face value doesn't seem to me to require that we attribute anything as elaborate as a recognition on the part of the whales of a community that includes both them and us.

This suggestion is roughly in the spirit of the proposal, by Robert Gordon and others, that simulating the other's state of mind may be the most practical way of coming to know another's mind (Gordon 1986; cf. Davies and Stone 1995). This proposal, while speculative, is not wholly without empirical foundation. There is evidence that such simulation might actually be mediated by specialized neurons, which -- whether as a result of learning or maturation -- are activated quite precisely by the observation of certain goal-related motor actions as well as while in the process of executing the same actions. These finding about "mirror neurons" have led to the surmise that we share a capacity for "intentional-state attribution" with non-human primates. Such attribution is something less than a full-fledged grasp of the mental states of others, yet it "constitutes a necessary phylogenetical state within the evolutionary path leading to the fully developed mind-reading abilities of human beings." (Gallese and Goldman 1998, 500). There is nothing in this research that applies specifically to emotions; but it doesn't seem far-fetched to extend the speculation to the apprehension of others' emotional states. To give it at least some plausibility, it may suffice to point out that expressions of emotions involve motor activity, and that among the fascinating findings of Paul Ekman and his team of researchers into emotional expression, is the fact that a configuration of facial muscles brought about by learned direct control of those individual facial muscles, when it fully coincides with the expression of a typical emotion, is liable to induce the feeling of that emotion (Ekman and Friesen 1975) (Ekman 1999).

In the light of these considerations, if we think of compassion as a rather more primitive capacity for emotional cognition, we might even be able to agree with Nussbaum that compassion is "the basic social emotion", while side-stepping, to some extent, the debate about the "evolution of altruism." For the capacity for compassion might actually be a side-effect of an obviously adaptive capacity to read minds, or what S.J. Gould has called a "spandrel". It's obvious that taking an accurate reading of the state of mind of someone who may be in a position to influence one's own well-being is a crucial advantage. But it may also be the case that this simply can't be done efficiently without some measure of empathy. To agree with this doesn't actually require that one settle the question of what this would consist in: ("Does one actually think... that one is the sufferer? Does one imagine one's own responses as fused in some mysterious way with those of the sufferer?" (Nussbaum 1996, 34). All that is required is that one's apprehension of the state of the other's mind be accurate, and that this very accuracy carry as a side-effect the power to move the mind-reader. This side-effect may sometimes be a costly one, if it leads to altruistic behaviour, but cannot be so costly as to outweigh the general advantage it provides.

The process by which such a capacity to "read" another's mind emotionally evolved might exemplify the intriguing mechanism known as the "Baldwin effect." (Baldwin 1896, Hall 2001, Deacon 1997, 322ff.) The Baldwin effect was originally conceived as a way to explain certain Lamarckian appearances in strictly Darwinian terms. In its most general form, it consists in a feedback loop that begins with non-instinctive behaviour, which brings about a change in environment, which change, in turn, leads to new selective pressures which favour genetic dispositions for certain types of behaviour. When the behaviour at the end of this loop is of the same type as the behaviour at the beginning, we may have something that looks rather like an episode of Lamarckian evolution, since it will actually be the case that a particular (group of) organisms' choosing a certain mode of behaviour can ultimately lead to a predisposition for that type of behaviour to be coded in the genes. But that happens only in special cases; in its full

generality, the Baldwin effect merely records the fact that behaviour can influence genes; the indirect character of the influence will usually guarantee that the results will be entirely unpredictable, so that what the Baldwin effect actually does it to lead the evolution of organisms capable of behaviour in wholly novel directions.

In the present case, what I am suggesting is that the sense of community might be not the precondition but the unpredictable distant effect of a change in our ancestors' mode of life which increased their mutual dependence. Here is how it might have worked: assume a change in the behaviour of a population which makes only a beginning of sociality. That given, an obvious advantage will accrue to those who are capable of 'mind-reading'. Under those circumstances, it may well be that the most efficient method of mind-reading is the one that enlists the emotions. (Otherwise, as Nussbaum rightly notes, we have failed to "distinguish between really accepting a proposition and simply mouthing the words"; thus Nussbaum comments, à propos of Rousseau's "To see it without feeling it is not to know it", that "this... means something very precise: that the suffering of others has not become a part of Emile's cognitive repertory...." (Nussbaum 1996, 38). That being the case, we reach a stage of much fuller sociality, a stage at which, as Nussbaum comments of the viewer of tragedy, one is brought by the compassion evoked by the tragedian's or the novelist's art to "see inequalities in vulnerability: therefore .... one will have reasons to raise the floor of security for all." (45) And in this way one's innate capacity for empathic compassion will have the effect of bringing us close to the stoic ideal of moral development of "moving the [concentric] circles [of concerns progressively more distant from our own purely egoistic ones] progressively closer and closer to the center, so that one's parents become like oneself, one's other relatives like one's parents, strangers like relatives, and so forth." (Nussbaum 1996, 48)

Note that as I have imagined this story, it presents us with the necessary transition from the realm of what I have called the black-and-white projection of our emotions to the full-colour version. The story begins in the simple instrumentality of reading other minds; and in telling it

my one and very minor difference from Nussbaum's own formulation has been to stress the possibility that compassion might actually start simple, so that we don't need to ask "How could such a complex sentiment be learned?" (Nussbaum 1996, 39) but rather "How could this simple and robust instinct have evolved?". But in order to yield the full image that such a reading implies, we need to see beyond the one-dimensional need to make choices: to be sure, my knowledge of other minds is first relevant to me for practical purposes, but if the emotional life is anything like the sort of thing that Nussbaum describes (and that I have tried to capture with the notion of paradigm scenarios), then to know another's mind is to know something that simply cannot be adequately rendered in black-and-white -- in the two-dimensional terms of choice to act or not to act. They constitute rich experiences of incommensurable dimensions of value, and they are what compassion or empathy actually strain to capture.

The Challenge of Nasty Emotions

Among the arguments that Nussbaum makes in defense of the moral relevance of pity, are some designed to parry the complaint that where pity is allowed, so will far less acceptable passions. One who accepts the values implicit in taking some things to be genuinely worthy of pity "accepts that she has given hostages to fortune. And to give hostages to fortune is to be set up not only for pity, but also for fear and anxiety and grief--and.... for anger and the retributive disposition as well." (Nussbaum 1996, 43). Nussbaum's own answer to this is that the champion of the moral significance of emotion "is not committed to saying that any and every damage, slight or insult is an appropriate occasion for retributive anger. By far the largest number of the social ills caused by revenge concern damages to fortune, status, power and honor, to which the defender of pity does not ascribe much worth."(Nussbaum 1996, 49). This answer seems to me plausible enough; but I would like to place it in a somewhat broader perspective.

It is to be expected that those who defend the importance of emotions to the rational and moral life should find themselves accused of ignoring the fact that some emotions are *nasty.* An emotion can be said to be nasty if one of the following obtains:

(i) to experience itis strongly aversive;

(ii)   it is exposed to strong disapproval; or

(iii)   it is always or usually followed by bad consequences.

Nasty emotions pose a challenge for the defender of emotions as perceptions of value. How can one square the axiological perspective with the existence of nasty emotions? Does the existence of nasty emotions have any tendency to undermine the view of emotions as apprehensions of axiological values?

The answer lies in attending to the difference between what I have called the black-and-white and the full-colour perspective on emotions. In a purely black-and-white world, that is, where the immediate practical dimension is the only one to be considered (a perspective in which, I have argued, the simplifications of immediate hedonism can actually bring a surprising amount of illumination), nasty emotions would be condemned simply in virtue of the fact that their existence has, on the whole, a negative influence on utility, whether directly or in combination with their consequences. Whether "damages to fortune, status, power and honor" are deemed important or not, they do constitute damage, and are the standard consequence of the emotions focused on revenge. But on the axiological view the issue isn't whether they have bad consequences, any more than it can be an epistemological question whether a given belief might have bad consequences. True beliefs can be discouraging or even depressing (there is also, notoriously, interesting evidence to suggest that depression can favour true beliefs (Pacini et al. 1998)); but that is not epistemologically relevant: the consequences of having a belief are not constitutive of its correctness. Similarly, having a "correct" desire, that is, a desire for something that is "good" or actually desirable, may have highly undesirable consequences: but the consequences of having the desire are not relevant to the question of whether the object of the desire was indeed as the desire represented it, that is, good or desirable. What determines whether a given state is "correct", in its own terms, is whether or not the target to which it is directed matches the state's formal object. In the case of belief, the analogous dimension is truth;

in the case of desire, it is goodness or desirability. In the case of emotion, that is determined by its defining paradigm scenario: that formal object lays out an axiological dimension. but we must resist the temptation to think there is some single dimension that constitute something like emotional appropriateness, for in truth each emotion has its own standards of appropriateness. It is a further problem, separable from any particular epistemological question about a given belief, whether that particular belief is or isn't a likely component of a good life in a particular context. Similarly there are, for each emotion, two separable questions, one of which concerns the extent to which the emotion is fitting to its object in terms of its object's relation to the emotion's formal object, and the other concerns the contribution that such an emotion, in the present context, might make to the good life. What is needed here is what I've called elsewhere axiological holism (de Sousa 2000). In our full-colour world of indefinite emotional dimensionality, we can judge the value of a particular emotion only in the comprehensive context in which it is experienced. Each emotion has arisen in a specific way, to form its own specific formal object; but there is nothing in either natural selection or individual biography to provide for any sort of intrinsic teleology, to dictate that all emotions will be useful, pleasant, altruistic, or in any other way commendable. Nevertheless, its very existence makes some claim on us, to constitute some part of a life someone might want to live. That claim can well be outweighed by competing claims; and this may be so even where the emotion is generally speaking one that fits in with a reasonable conception of the good life. Socially useful emotions such as compassion or benevolence might on some occasions mislead us, just as the undoubted utility of our faculty of vision does not guarantee that visual illusions never occur. That emotions in general constitute apprehensions of axiological reality does not imply that every emotion is equally trustworthy. This holds at both the levels of appraisal just distinguished: some emotions may be misleading in their context while being fitting in themselves; others may be systematically misleading or nasty. In either case, we might speak of axiological errors. But that doesn't show that there isn't something that each emotion gets right; nor does it show that an emotion's getting something right is sufficient to guarantee that the emotion is a good thing.

One more complication must be confronted. Though many philosophers have been inclined to think of emotions as products more of nature than of nurture, I believe most would agree that the emotions most important for moral development are learned. This might be partially true even for the universal, "basic" emotions; but it's at least surely true of all the others. That doesn't necessarily mean that they are uniform throughout any cultural group; but it does mean that they arise either in conformity with group pressure or in reaction against it: either way, a child's emotional repertoire is hardly likely to escape the cultural influences that have shaped it, even if a few vectors get to change signs. The iron rule of most societies is conformity. This is made very clear in Sober and Wilson's survey of cultures of the world: societies such as ours are exceedingly rare, where the individual is not forced by the tyranny of the group to conform to social expectations, especially in everything that might remotely affect reproduction, kinship, and sexuality. (Sober and Wilson, 1998, chapter 5). Contemporary examples are all too easily available. Not long ago I witnessed an interview of several men who had killed a sister or daughter on a matter of "honour" -- that is, because the victim had had been raped. The most astonishing thing about these men's attitudes is that they appeared not to suffer the slightest twinge of ambivalence, let alone guilt, about what their sense of "honour" had moved them to do. Such is the power of social conditioning.

The moral is that axiological holism can't mean merely that we must look beyond the individual's emotions to those endorsed by the society; for by our lights many of the emotions endorsed by most societies are unspeakably vile. We can judge whether that emotion's contribution is right or wrong only in much in the same way as we test the veracity of perceptual information: by looking at the same thing from different points of view, which means, in effect, by appealing to other perceptions or, mutatis mutandis, to other emotions. Something like the method of reflective equilibrium is commonplace in science as well as in ethics; what is less often noticed, and what Nussbaum's work in particular has made abundantly clear, is that in the case of ethics, the items that need to come to equilibrium are typically not ordinary empirical

facts, but emotional responses. Those emotional responses are best understood as axiological perceptions; and to take full account of them it is not sufficient to attend to their two-dimensional projections onto the space of pleasure and pain, or even the half-dozen dimensions afforded by the "standard" emotions. The lower-dimensional projections of a higher-dimensional world may be not merely misleading, but positively incoherent, just as the two-dimensional drawings of Escher represent something impossible in the three dimensions which they purport to depict. That is why, in a full-colour world of real emotions, the high degree of specificity entailed by art and literature "fully imagined" are central, not merely auxiliary, to the task of moral and emotional learning.

References

Baldwin, J. M. 1896. Consciousness and evolution. *Psychological Review* 3:300-08.

Campbell, S. 1998. *Interpreting the Personal: Expression and the Formation of Feeling*. Ithaca: Cornell University.

Clynes, M. 1977. *Sentics: The Touch of Emotions*. Garden City: Anchor Press.

D'Arms, J., and D. Jacobson. Forthcoming. The Moralistic fallacy. *Philosophy and Phenomenological Research.*

Darwin, Charles. 1998 <1872,1889>. Ed. with intro, afterword and commentaries by Paul Ekman. *The Expression of the Emotions in Man and Animals.* London: Harper Collins.

Davies, M., and T. Stone, eds. 1995. *Mental Simulation: Evaluations and Applications*. Readings in Mind and Language. Oxford: Blackwell.

de Sousa, R. 2000. Learning to be natural. In *Being Humans*, ed. Neil Roughley. Berlin and New York: De Gruyter. 287-307.

------. 1974. The good and the true. *Mind* 83:534-51.

------. 1987. *The Rationality of Emotion*. MIT Press <imprint: a Bradford Book>.

Deacon, T. W. 1997. *The Symbolic Species: The Coevolution of Language and the Brain.* New York: Norton.

Ekman, P. 1999. Emotional and conversational nonverbal signals. Conference presentation. San Sebastian.

Ekman, P., and W. Friesen. 1975. *Unmasking the Face: A Guide to Recognizing Emotions from Facial Expressions*. Englewood Cliffs, NJ: Prentice-Hall.

Gallese, V., and A. Goldman. 1998. Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences* 2(12):493-501.

Gallop, D. 1960. True and false pleasures. *Philosophical Quarterly.*

Gordon, R. M. 1986. Folk psychology and simulation. *Mind and Language* 1:158-71.

Gould, S. J., and E. Vrba. 1982. *Paleobiology* 8:4-15*Exaptation: A Missing Term in the Science of Form.*

Gould, S. J. 1979. *Proceedings of the Royal Society of London* 205(581-598)*The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme.*

Griffiths, P. E. 1996. The historical turn in the study of adaptation. *British Journal for the Philosophy of Science* 47:511-32.

Hall, Brian K. 2001. " Organic selection: proximate environmental effects on the evolution of morphology and behaviour". *Biology and Philosophy* 16: 215-237.

Humphrey, N. 1984. Nature's psychologists. In *Consciousness Regained: Chapters in the Development of Mind*, 29-41. Oxford: Oxford University Press.

Itard, J. 1964. Mémoire et rapport sur victor de l'Aveyron. L. Malson. In *Les Enfants Sauvages: Mythe et Realité*. Paris: Union Générale d'éditions.

Masson, J. M. 1995. *When Elephants Weep: The Emotional Lives of Animals*. New York: Delacorte Press.

Nozick, R. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.

Nussbaum, M. 1990. *Love's Knowledge*. Oxford, New York, Toronto: Oxford University Press.

------. 1994. *The Therapy of Desire: Theory and Practice in Hellenistic Ethics*. Princeton: Princeton University Press.

------. 1996. Compassion: The basic social emotion. *Social Philosophy and Policy* 13:27-58.

------. 1997. *Cultivating humanity : a classical defense of reform in liberal education*. Cambridge, MA: Harvard University Press.

Pacini, Rosemary; Muir, Francisco; Epstein, Seymour, 1998. Depressive realism from the perspective of cognitive-experiential self-theory. *Journal of Personality and Social Psychology.* 74: 1056-1068

Penner, T. 1970. False anticipatory pleasures in the Philebus 3613-41a6. *Phronesis*:166-78.

Rosenfield, I. 1992. *The Strange, Familiar and Forgotten*. New York: Alfred A. Knopf.

Sober, E., and D. S. Wilson. 1998. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.