

Destructive emotions

Owen Flanagan
Duke University

This paper discusses the problem of destructive emotions by comparing Eastern and Western assumptions about emotions. In the case of anger, for example, Eastern thinkers straightforwardly posit that it is entirely possible to cultivate attitudes in which anger is naturally absent. In the West, by contrast, it is generally assumed that anger is a “basic” emotion that can be suppressed or managed, but not eliminated from one’s basic emotional constitution. Thus, in the Eastern way of thinking, emotion is a force that more easily harmonizes with rational approaches to life and to the specific problems in life.

Keywords: emotion, mind, mindfulness, Buddhism, anger

Introduction

My paper takes up the topic of destructive emotions, or better, destructive states of mind that have at their core a strong emotional component. The major issue that concerns me is the question of whether and, if so, how or under what circumstances, our emotions might be destructive or in, some sense, harmful to the agent who experiences the emotion or to the others at whom they are aimed or expressed — self-destructive or harmful to others. Whether and to what degree an emotion is, or might be, conceived as destructive depends, in some measure, on certain broadly empirical information such as whether the emotion is a biological adaptation, how it normally feels, what its normal psychobiological and behavioral features are, what its natural and original, as well as its current intrapersonal and interpersonal functions are, and so on.

But whether an emotion is destructive — even a basic, natural and original emotion — also depends, probably mostly, on the *normative* picture we have of

a person, a good life, one that conduces to human flourishing. We are cultured creatures and almost all the emotions, even the ones that are excellent candidates for being original and natural, and adaptations to boot, are now *extended* traits; that is, they are activated by situations that are very different from the situations in which they were originally designed to do their fitness-enhancing jobs. It is one thing to be scared by a fierce animal or to be angry with a stranger who threatens me bodily harm. It is quite another to be scared that I will not get a big raise — suppose my credit card debt is high — and to be angry with the Dean for not knowing or caring about this. Furthermore, and relatedly, what conduces to *biological fitness* and what conduces to *flourishing* are different things. Although, if we are lucky — and I think we are to some extent — Mother Nature may have given us emotions that are not intrinsically counter-moral, not completely at odds with flourishing.

A destructive emotion can have positive or negative valence. Emotions such as anger and fear have negative subjective valence. Love, on the other hand, normally has positive valence. Certainly a state of mania is destructive, on most every view. If this is right then certain emotions can feel good but be harmful or, at least, in tension with morality. Some think that even love, especially romantic love, is destructive at least in some respects. Within both Eastern and Western moral theory there is controversy over whether love impedes impartiality. It is not remotely odd to ask whether romantic love, love for one's children, or special love for one's friends, despite being good, might be in tension with ethical impartiality or universal compassion.

In this paper, I am particularly concerned with what I will call the “moral emotions”, specifically with what P. F. Strawson calls the “reactive attitudes” in a famous 1962 paper called “Freedom and Resentment”. One reason for raising the topic is this. Strawson's claim is that the reactive attitudes are “original and natural” in the way inductive reasoning is, something we could not ever give up. Luckily, the reactive attitudes — if suitably expressed — are good. They are attitudes through which we express our common humanity, recognizing others as persons, and according them the positive and negative expressions that come from seeing and treating them as moral agents.

Within Tibetan Buddhism, however, the emotions, especially anger — but many others as well — are thought to be bad, something we should give up. In the West we believe “ought implies can.” His Holiness the 14th Dalai Lama tells me this is believed in the East as well. Thus the belief that anger ought to be eliminated is not an idle, impossible ideal within Tibetan Buddhism. It is thought to be possible. Indeed, it is thought to be a realistic goal that many

Buddhist practitioners have achieved or are on the way to achieving.

There seems this asymmetry between thinking inductively and feeling and expressing the reactive attitudes: No one thinks it is advisable, even if it were possible — which it is not — to stop thinking inductively. But some, Tibetan Buddhists, for example, think it possible and advisable to overcome, even eliminate what is “original and natural” in the case of the emotions. If this is possible, “overcoming the genotype” let us call it, that alone is interesting. We have no practices in the West as far as I know, save medical ones, to overcome completely the genotype. We do have practices, just as we do for induction, to moderate, modify, and adjust our emotions, our reactive attitudes — to make our emotions “apt.” And we engage in such practices in accordance with some conception of what is good and bad, constructive and destructive. Tibetan Buddhists, however, seek *elimination*, at least for certain emotions, in accordance with their conception of the good. I’ll be saying some things that are relevant to the question of whether this is considered possible and desirable from the point of view of Western moral philosophy. Hopefully my remarks will generate some helpful thoughts about whether and to what degree Tibetan Buddhism is onto something — that is, whether we would, if we only knew how — think it best to go farther down the road of emotional elimination. I think — in fact, I know — that most of us find this idea distasteful. It is good to think about why that is. I won’t say much directly in this paper about this last point, that is, I won’t offer a diagnosis of why the path of emotional elimination is thought distasteful in the West. But I do hope to engage the minds of my readers in that question.

The reactive attitudes

The reactive attitudes comprise the set of human responses that include indignation, resentment, gratitude, approbation, guilt, shame, pride, hurt feelings, feelings of affection and love, and forgiveness. Strawson’s interest in the reactive attitudes came from the fact that the reactive attitudes are centrally involved in morality, express our humanity, and are underestimated in a variety of ways by what he called “intellectualist” moral theories.

Strawson claims that (1) the reactive attitudes are part of the normal and original conative repertoire of members of the species *Homo sapiens*; (2) the reactive attitudes *express* normal human reactions to acts, traits, dispositions, or to whole persons; (3) the normal expression of the reactive emotions involves interpersonal relations where benevolence or malevolence is displayed or, at least, where they are at stake.

The last point might make it seem as if the reactive attitudes are invariably other-regarding. This is a mistake. We can, indeed we often do, experience and direct reactive attitudes towards ourselves. Reactive attitudes are thus sometimes, possibly often, self-regarding. A reactive attitude — pride, guilt, shame, feeling one is not getting one's due or a feeling of obligation — can be experienced first-personally in virtue of direct social feedback, for example, through the expression of indignation or approval on the part of another. It can even be self-applied, as one simply contemplates some act or feels disposed to act in a certain way, without in fact actually (now or ever) carrying through the action.

So, the reactive attitudes are self-regarding as well as other regarding. They are also experienced and expressed vicariously, that is, on behalf of others. Vicarious expression of the reactive attitudes occurs even in cases where there is no actual impending good or bad will directed towards us as an individual. How widely our vicarious expressions of reactive attitudes spread is a variable matter, probably highly dependent on moral education and culture. What we do know is that they can spread or extend very widely over space and time. Contemporary persons, with knowledge of human history or simply through exposure to television news, do feel vicarious indignation towards acts of malevolence of people who are long dead or engage in atrocious practices half way around the world, and who in neither case present any reasonable threat to us, our loved ones, or our countrymen.

Now Strawson doesn't explicitly state that the reactive attitudes come with our natures, or better, he does not explicitly state that they are a product of our evolution as a certain kind of animal. But it is certainly not unfair to attribute to him some such view. Strawson emphasizes that these attitudes are natural, that they appear in some form across all cultures, and that their ubiquity has something to do with our nature as social creatures.

The evolution of the reactive attitudes

One reason for thinking that Strawson would approve of thinking of the reactive attitudes as products of evolution is that in a telling footnote he compares the reactive attitudes to induction.

Strawson writes: "Compare the question of the *justification* of induction. The human commitment to inductive belief-formation is original, natural, and non-rational (not *irrational*), in no way something we choose or could give up." The idea, I take it, is to suggest that what goes for induction also goes for the reactive attitudes — they are "original, natural, non-rational (not *irrational*)...[and not something we ever] could give up."

Now, although I said at the start that we have no practices in the West to eliminate emotions, we do have theories about the rationality of the emotions, and we do have views about moderating and modifying certain emotions by Reason. One brand of intellectualism, especially common among compatibilists about free will and consequentialists about morality, recommends taking what Strawson called “the objective attitude.” The basic idea is to cognitively penetrate our natural and original conative economy and not to take benevolence or malevolence too personally, indeed as far as it is possible not to take these things personally at all. The reason is simple: she who does you good or harm could not have done otherwise. From a rational point of view, the only purpose to be gained from displays of positive or negative reactive attitudes involve manipulating the behavior of others in the future. Likewise, first personal regret and remorse are irrational if one has done one’s best, or alternatively they are ok, to a point, if they will keep one on track down the road.

So we do have views in the West that deem certain emotions irrational, and to that extent not apt. But to the best of my knowledge, no one who has proposed taking the “objective attitude” has believed we could actually pull it off. In any case, if we take Strawson’s point about the reactive attitudes being “original and natural” to be a concession to the plausible idea that we evolved to express the reactive attitudes in the relevant or appropriate interpersonal or intrapersonal situations, then several questions arise:

1. Did the reactive attitudes arise as *adaptations*?
2. If so, are they still adaptive?
3. How *modifiable* are these attitudes?
4. If they are modifiable, are there any reasons to work to modify, moderate, or otherwise adjust how, when, and under what circumstances we experience or express the reactive attitudes?

The last two questions are especially important insofar as destructive emotions are on center stage. The reason is that if a trait is destructive we have reason to want to eliminate, change, or modify it insofar as is possible. But in the footnote where Strawson compares the reactive attitudes to inductive information-processing, which is original and natural, he gives reason for pause about how modifiable the reactive attitudes are. To remind you, he adds first that induction (the same goes for the reactive attitudes) is “*in no way something we choose or could give up*” (my italics); and then second, immediately after he says this — but here speaking only of induction: “*Yet rational criticism and reflection can refine standards and their application, supply ‘rules for judging of cause and effect.’*”

These comments might make us worry about the modifiability, or degree of modifiability, of the reactive attitudes, and especially about the prospects for “overcoming” in any strong sense our emotions, especially emotions that are natural and original. Addressing this concern is crucial for our purposes since in the West it is commonly assumed that *Reason* (the basic idea goes back at least to Plato) can, to some extent reign in excessive, destructive, or untamed emotions. And in the East there are both reasons and practices in place to work on modifying, moderating, eliminating, or even to continue experiencing but ceasing to identify with, negative or destructive emotions.

What an adaptation is

There has been of late some fun fighting in the streets, or at least in the pages of *The New York Review of Books*— one of the contemporary intellectuals’ favorite sites for engaging in their preferred version of street fighting or pub brawling — about the issue of adaptation and how committed one ought to be the program of adaptationism. Stephen Jay Gould accused Daniel Dennett of being an “ultra-Darwinian fundamentalist”, of committing the sin or crime of “pan-adaptationism.” A panadaptationist, dubbed elsewhere by Gould and Lewontin, to be a “Panglossian”, allegedly thinks that Mother Nature produces all and only adaptations. That is, Mother Nature produces all and only traits that contribute to inclusive genetic fitness, as inclusive genetic fitness, formerly known as reproductive success is now understood within contemporary population genetics. An associated sin, crime, or misdemeanor commonly committed by an ultra-Darwinian fundamentalist is to think that all traits that evolved via evolutionary processes are optimal. The familiar response to optimific thinking is to point out that Mother Nature tinkers with designs already in place, designs that constrain what new traits she can produce.

Karen Neander, a philosopher of biology, expanding on an example from Gould writes that: The Panda’s “thumb” was selected for stripping leaves off bamboo and that therefore is its teleonomic function. But it does not follow that the Panda has the best bamboo-leaf stripper it could possible have, or that natural selection worked toward an adaptive outcome in this case in the absence of all constraints. Far from it (Gould, 1980). While the Panda’s thumb has a clear teleonomic function, it is a wonderful illustration of the fact that natural selection is a tinkerer and a satisficer, heavily constrained by the past and by the alternatives that are presently available. The Panda’s thumb — in fact, an elongated wrist bone — is an imperfect design from an “engineering point of view.”

The point is that even adaptations such as the Panda's thumb, in virtue of being constrained by pre-existing developmental constraints, are from an engineering point of view *satisfactory*, but not necessarily *optimal*. Furthermore, and contrary to ultra-Darwinian fundamentalism (if there really is such a sect), not all traits produced by evolutionary processes are adaptations.

Everything we can do is allowed by our genes, but not every capacity was selected *for* because it was an adaptation. Elliot Sober draws a useful distinction between selection *of* and selection *for* a trait. It is a good that our inductive capacities were selected *for* because they led to success in locating and relocating food sources, suitable shelter, mates, and the like. It turns out that these inductive capacities that were selected *for* allow us, with certain refinements acquired in cultural contexts, to do at least elementary science. The ability to do basic science was not however the reason our inductive capacities were selected *for*. And this is because, at least in the original evolutionary context, displaying these abilities (a) did not occur and (b) therefore did not contribute causally to our inductive capacities being selected *for*. The inductive capacities required for doing basic science piggybacked on the inductive capacities that led to success at hunting, foraging, and mating. There was selection *of* the relevant capacities to do science, but not selection *for* doing science.¹

It is standard practice in evolutionary biology to restrict talk of adaptations to traits that were selected *for*. In the case of induction, the enhanced inductive abilities humans displayed when they got around to doing what we now call 'science', were not selected *for* to do science in the originally evolutionary context. Cultural selection and learning, and the development of certain social practices, played a big role here, a big enough role to resist calling doing science, certainly to resist calling the deployment of the canons of inductive logic, statistics, and probability theory, as we now utilize them in doing science, as capacities that were selected *for*, and thus to refuse thinking of the display of these traits as a set of adaptations.

The moral emotions, again

Brains, notoriously, have left us a pathetic, virtually nonexistent, fossil record of how they were used by the humans in which they were initially housed. In making our surmises about our natures when we evolved, we depend heavily on remnants of primitive social groups, on tools, evidence of fire use, skeletal remains, knowledge of food resources together with DNA testing of contents of abdominal cavities to see whether eating one's fellows was rare or common, as well as knowledge of geography, weather, migration patterns, etc. In any case,

if we look at the phylogenetic record, especially at data gathered about other species of hominids, as well as observations of extant species with common ancestors — chimps, bonobos, and great apes, for example, we get a better fix on whether a trait we display is or might be an adaptation.

Putting these data sources together, and following the insights provided by Darwin in *The Expression of the Emotions in Man and Animals* (1872), as well as by the important work of Paul Ekman and his colleagues a century later, it seems safe to say that certain human emotions are universal. There are several reasons for saying this: first, homologues appear in other animals — canines, as well as close ancestors — of the emotions we experience and express. Second, in social species there are characteristic movements of the facial musculature that are recognized for what they are (e.g., for the behavioral dispositions they display) by conspecifics who then seem to respond appropriately to the particular display. Third, for the basic emotions of *fear, anger, surprise, happiness, sadness, disgust, and contempt* we have or are well on our way to locating better and better physiological markers that distinguish among the different emotional expressions. Fourth, the emotions, or at least the relevant facial expressions, alleged to be universal are in fact recognized across human societies — among pre-literate New Guineans as well as native New Yorkers — for what they are.

Let us accept therefore that there are universal human emotions that have evolved to be essential traits of the members of the species *Homo sapiens*. They, the basic emotions, are part of the original equipment we enter the world with, just as eyes, ears, noses, and noisy hearts are. But now several questions arise.

First, there is the question of whether Strawson's reactive emotions are basic in the way Ekman's emotions are. This is a complicated issue. Guilt, shame, indignation, gratitude, and resentment are not on Ekman's list, but they are on Strawson's. One obvious move to make here, assuming one wants to accept Strawson's idea that the reactive attitudes are natural and original, is to link them as closely as possible with the emotions on Ekman's list. One can get a fair distance with this strategy. So, for example, we can say that indignation is a type of anger directed toward malevolent conspecifics; gratitude is one way we express happiness at being well-treated; sadness is linked with happiness insofar as it is its inverse. Sadness is an emotion we express in virtue of losing rather than attaining or retaining some good, say, some thing or some relationship we desire. Guilt, possibly shame as well, might be thought of as anger (or in the case of shame, possibly surprise) directed inward. Relatedly, feelings of regret and remorse might be seen as tied to feelings of real or impending loss, and again to anger directed towards oneself for doing something that does or might

bring about such a loss. I'm not going to fuss over how in detail the entire mapping would work. I will simply assert that it seems possible to draw some fairly close links between Ekman's basic emotions and Strawson's reactive attitudes, and to make the case that they are both homing in on similar, if not exactly the same set of phenomena, as they draw up their lists of that part of our conative economy that is original and natural — that comes with being human. Sometimes, Strawson (showing his Kantianism) speaks of the reactive attitudes as being activated when respect for my person fails to be acknowledged, or when my dignity is challenged. I doubt that our ancestors, who involved tens of thousands of years ago at the end of the Pleistocene as the ice melted, conceived of themselves as suffering challenges to their dignity and self-respect. We Westerners, we children of the Enlightenment, do sometimes, possibly often, experience affronts to our dignity as persons, and feel as if we are not treated with proper respect. Strawson can make room for this difference between the reactive attitudes of ancestral and modern persons. Although when it comes to the reactive attitudes Strawson is not a social constructionist — remember the reactive attitudes are natural and original — he is clear, in a way no one else, to the best of my knowledge, has pointed out, that there are individual and cultural differences in everything from the intensity of the experience or display of each of the reactive attitudes to many of the conditions that activate the reactive attitudes. Getting a haircut the day after one's father dies is not something an orthodox Hindu ought to do — it comes close to sacrilege. For us getting a haircut is a perfectly reasonable thing to do in preparation for any funeral. The fact remains that in both cultures there are ways of behaving at times of the death of a close relative that will elicit disapproval.

In any case, if I am right that the set of reactive emotions lies in close proximity to, or can be plausibly understood as variations on, or species of certain of Ekman's basic emotions, then there is an interesting consequence in the offing. It is this: the basic emotions are moral to some degree, or at least they are *proto-moral*. In saying the emotions in question are 'moral' or 'proto-moral', I mean simply to claim that they appear to have been designed and thus to function to express positive or negative feeling, typically where such feelings are elicited by interactions with others. One might object that for an emotion to be, or to function as a *bona fide* moral emotion that there must be a moral community in place, and a conception of the demands made by that community held in the minds of its members. I don't want to fuss over this issue, partly because I don't think too much turns on it. If the word 'moral' causes discomfort, call the emotions in question 'proto-moral'. All that is being asserted by

calling them proto-moral is that the individuals involved in social interaction (there is, to be sure, the assumption that we are and always have been social animals) prefer to have certain experiences and not to have certain others, that there are goods they desire and bad things they dislike, and that purveyors of good or bad things will be the objects of positive or negative reactive attitudes.

We can safely assume that the reactive attitudes function in the original evolutionary situation to convey feelings about interpersonal commerce. We need not take a stand on whether they are, at that point in time, legitimately dubbed moral emotions, or on whether deep-down-inside we are all rational egoists, concerned only for our own good, or whether we are endowed with a heavy dose of benevolence and fellow-feeling, as well as a selfish streak. It is enough to imagine us involved in social relations with certain likes and dislikes, living among compatriots who either inadvertently or on purpose do some things that we like and some things that we don't like.

Now it would keep certain things neat and clean if the proto-moral emotions were adaptations. The prospects that they might be adaptations appear *prima facie* promising. First, attitudes such as anger and fear are easy to imagine having credible links to fitness. Closely related species that display these attitudes commonly do so when they are in physical danger. Second, attitudes such as happiness and sadness might well subserve various types of social interaction that lead to fitness. If I find sex pleasant I will mate, and I will convey my pleasure, my happiness, possibly gratitude, to whomever it is I find it pleasant to mate with. If I find playing or being with others pleasant I will be concerned that they fare well. Dispositions to express reactive attitudes vicariously might easily be imagined to naturally extend to my offspring or to my mate, as well as to any others whose company I find pleasant, who do me good, or who I see are in a position to do me good. Likewise, assuming I have come to care about certain others (even if for totally selfish reasons, as we say) I will be disposed to experience anger towards those who do, or seem likely to, harm them, possibly even to act on this anger. Indeed, primatologists often speak of "moralistic aggression" among nonhuman primates where the aggressive display can arise when a chimp is directly threatened or when another chimp he cares about is. Finally, an early warning system where we display our emotions facially before we act on them seems like a good design strategy for creatures who should, all else being equal, wish to make their feelings and desires known without being maimed or killed.

These are some of the things that can be said for the idea that the proto-moral emotions, those comprising some relatively large subset of Ekman's set

of universal basic emotions as well as Strawson's natural and original reactive attitudes, are adaptations in the biological sense. But there are two caveats.

First, evolutionary biologists and philosophers of biology standardly make a distinction between the question of whether or not a trait is an adaptation in the strict sense, this being largely a matter of history, in particular a question of whether the trait was fitness enhancing when it evolved, and the different question of whether it is adaptive in current environments, where the relevant sense of 'adaptive' is tied to current fitness-enhancing features. This is a difference that makes a difference about which I will say more shortly. Second, I have said that there seems to be a pretty straightforward way to read Ekman and Strawson as working the same terrain, and thus to understand our conative economy as *naturally structured for morality*. The idea is that our emotions, or an important subset of them at any rate, were designed for the purpose of suiting us for interpersonal commerce, given that fitness stands or falls on how well we do in such commerce. This is tantamount to saying that the moral emotions, or the proto-moral emotions, are a product of selection-*for*, not simply the result of selection-*of*. The moral emotions causally contributed to fitness among early humans, possibly because homologues of them had been selected for in ancestral species from which we descended. The proto-moral emotions did not piggyback in as a neutral side effect of some other trait, say the ability to dig for nourishing roots. No, according to the story I am telling, the proto-moral emotions are not only natural and original, they were selected *for* to serve the inter- and intra-personal functions they in fact serve. Furthermore, since we are speaking at this juncture about biological not cultural evolution, the proto-moral emotions are not, or should not, if we have identified them correctly, turn out to have a heavy cultural overlay. Let me take each point in turn.

First, suppose we take it for granted that the proto-emotions, or some subset of them, can be plausibly defended as adaptations according to the criterion that weights most heavily the causal contribution of a trait to fitness in the original evolutionary situation in which the trait evolved and proliferated. It does not follow that the trait is now adaptive, where the meaning of adaptive is tied to being fitness-producing or -enhancing now.

The leaves of eucalyptus trees evolved to be glossy. This serves to preserve moisture, which is a good thing for a tree to do if it fought for survival in an arid climate. If, however, the climate in Australia changes to a tropical one, then the adaptation subserving moisture retention is no longer fitness-enhancing, and eucalyptus trees will become oversaturated with retained water and rot

away. Eventually there will be no more eucalyptus trees and no more koalas bears that depend on them, etc.

The relevance of the example to the case of the proto-moral emotions is straightforward. Even if these emotions were selected for and maintained in the species because they were adaptations in the original evolutionary situation (either our own or that of ancestral species from which we descended), this does not establish that they are adaptations (or adaptive) in the environmental niches we now occupy. This is one reason I simply disagree with Ekman when he says (personal communication) that the idea of destructive emotions is an oxymoron. His guiding idea seems to be that the basic emotions are adaptations and an adaptation cannot be destructive. What is true is that an adaptation cannot fail to be fitness-enhancing at the time it evolves. But as with the eucalyptus trees I just imagined growing in a radically changed environment an adaptation can cease to be fitness enhancing if the environment changes enough. The reactive attitudes and the basic emotions may well have evolved as adaptations in close ancestors or in us (it makes a difference which it was, but I don't have space to go into the matter here). Knowing whether they continue to function as such depends on better understanding than we now possess of relevant differences between the original evolutionary contexts in which they arose and the environments we now live in. But the main point is that any trait can have been constructive, served a proper evolutionary function when it was selected for originally, but cease to do so in current environments if those environments are different enough. In that case, we can say, and say rightly, that the trait is now destructive from a fitness-enhancing point of view.

Second, there is another way — actually two other ways — an emotion can be destructive that doesn't depend at all on whether or which sense of adaptation we are using.

Ekman himself divides the basic emotions into positive ones, happiness, for example, and negative ones, anger or fear, for example. A negative emotion has negative qualitative feel, negative valence. Negative emotions are ones we don't want to experience and don't enjoy if we do. In this sense, they produce some measure of harm to us, even if overall they are fitness-enhancing. Fear and anger feel bad, even if fear gets me to head for the hills when danger lurks, and even if anger disposes me to defend myself from impending harm. In the best of all possible worlds, we would prefer not to experience or to have occasion to experience these emotions, and the reason is that they don't feel good. They take us out of some sort of comfort zone we'd rightly prefer to be in or stay in. It is common wisdom that harboring anger or fear is destructive to him or her

who harbors it. But I am willing to claim that simply experiencing fear or anger over a short episode is destructive to him or her who experiences it, at least in the sense that it is qualitatively unpleasant and produces a sense of unease and disequilibrium. There is a second way in which the negative emotions are destructive beyond the unease and disequilibrium they cause for the person who experiences the negative emotion. This point about destructiveness is easier to make than the first one. It is simply that anger, indignation, disapproval, and threats of ostracism are designed to make the object of these emotions feel bad, back off, change their behavior, cease and desist, leave, and/or fear reprisal. The point of these reactive attitudes is to threaten to do harm. We intend for the object of these attitudes to experience fear, guilt, shame, to change his behavior, to apologize, or to risk fighting and punishment. If it is true that negative emotions are destructive in the sense that they are qualitatively unpleasant and produce unease, then it follows that any expression of anger is destructive to her who expresses it. With regard to the person at whom it is aimed, since it produces fear or anger which again is unpleasant and a source of unease, then it is also destructive simply in virtue of possessing these properties. I'm inclined for these reasons to say that negative proto-moral emotions are *intrinsically* destructive whether one is the subject venting a negative emotion or the recipient of such venting. All of which, by the way, is compatible with such displays being biological adaptations. The point is that in virtue of their being unpleasant, we can and do want the world to include as few opportunities as possible for giving expression to or being on the receiving end of negative emotions. Never wanting to have a certain type of experience, if possible, is normally taken as evidence that the experience is unpleasant, and thus, in some sense of the term, destructive. Furthermore, to make the second point as clear as possible, negative emotions are destructive in the *extrinsic* sense that they express an intention on the part of the one who expresses the emotion to harm the other and to do so, if need be, in ways that go beyond the deliberate one-on-one act of hurting the other by expressing the negative emotion. Thus there is some form of harm or destruction aimed at and accomplished by the expression of the negative reactive emotion itself, as well as typically an implication that the harm can or will get worse unless the object of the emotion adjusts or makes amends for her character, attitudes and/or behavior.

Cultural coloration

I have already acknowledged something Strawson himself concedes, namely that the reactive attitudes are subject to forces of cultural learning. Some of the

reactive emotions, indeed the very names he uses for all the reactive attitudes, reveal a certain amount of local color, a color one might expect from a famous philosopher, a great admirer of Kant, who was of the manner born, and is a knight of the British empire. To speak a language of indignation, approbation, disapprobation, resentment, and gratitude is to speak in a more distinct and less widely understood idiom than to speak of anger, fear, sadness, happiness and surprise. Nonetheless, I have tried to make the case that the attitudes Strawson has in view are often reasonable facsimiles of Ekman's basic emotions. But I want to emphasize that I may be overstating the case, at least with respect to certain reactive attitudes on Strawson's list. Take pride, for example. According to Strawson, pride is a self-regarding attitude one has when one recognizes or feels that one has done what one ought to do, when one has done one's duty.

Now feeling pride in this sense, it seems to me, requires moving pretty far beyond the original evolutionary context in which I expressed willingness to call the basic emotions, or original and natural reactive emotions, proto-moral. Recall that I backed off insisting on calling them *moral emotions* because of the objection that for such emotions to be legitimately called 'moral' some conception of morality, even the institution of morality — some set of publicly recognized rules and principles — must be in place. I simply don't think it plausible to insist that the institution of morality was in place at the start, when *Homo sapiens* first appeared on the scene. Now an attitude such as pride, as opposed to simple happiness, seems especially to require that the institution of morality be in place and that a certain conception of a person, and in particular, what it means to be a good person, has developed. Furthermore, it is worth emphasizing that the normative notions of a person, and in particular of a good person or good persons, are different from, possibly orthogonal to the normative notion of a maximally fit individual. At eulogies when we remember a person for having been good, we almost never focus on their reproductive fitness, and this despite the fact that we might well talk about their deep and abiding love for their children, if they happen to have had any.

In any case, any self-regarding pride I may experience for doing what I ought to do, or what is my duty, may require as I have just said, that a certain conception of morality has been created or discovered. The same holds true, as I suggested earlier, for experiences of affronts to my dignity or feelings that I am not receiving the respect I deserve as a person.

Locke, famously, in his discussion of identity distinguished between the identity of a man, (in the generic sense of a member of the species *Homo sapiens*) and the identity of a person. Picking out an individual as a man — as

a conspecific — is pretty much a straightforward empirical matter and the term ‘man’ is straightforwardly descriptive. But the concept of a person is normative through and through. Adopting Locke’s point for present purposes, we might easily imagine our ancestors seeing each other as men, as fellow humans, without having any thoughts whatsoever of themselves or their conspecifics as persons, as creatures worthy of moral respect, worthy of recognition for their dignity, as possessing certain rights, and of themselves and others as having certain obligations, as agents who are responsible for their actions, and so on. To be sure the proto-moral attitudes in a situation of interpersonal commerce would no doubt have been implied, or been structured *as if* such a conception was in place, but I seriously doubt that it was in place in the original situation.

It may well have been that the *as if* situation gradually was recognized as being suited to the imposition of a certain moral conception and thus that the institution of morality and the normative notion of a person arose in this way. But that is as far as one can plausibly push an evolutionary account. The moral conceptions that arose in various places over the face of the earth, or the normative conception of a person fitted to the moral conception in question, or, for that matter, more complex and sophisticated reactive attitudes such as moral pride, respect for persons, and recognition of human dignity cannot have been original and natural in the sense that they were biological adaptations, traits that were selected *for* because they were heritable and causally contributed to fitness.

The analogy with induction is useful here. I have said that it is plausible that some sort of inductive capacities were selected *for* because such capacities causally contributed to fitness, to success at hunting, foraging, and mating. Suppose that there was selection for the straight rule of induction: If it is observed that regularity R occurs m/n of the time, infer that it will occur in the future to m/n . It is a familiar fact that this rule works fairly well in elementary situations, but it leads in more complex situations to bad reasoning. Ancestral humans might not have often confronted the situations I have in mind when social groups were relatively small, conspecifics were all well-known, and hunting and foraging ranged over relatively close distances. The bad reasoning I have in mind involves such fallacies as occur when sample size is too small or, what is different, when the sample is unrepresentative. The canons of statistics that include norms governing sample size and representativeness are now fairly well known — at least they are taught in methods courses at most universities. As Reichenbach pointed out these norms themselves were discovered — and it took a long time — by applying the straight rule of induction. The straight rule

was selected *for* and by deploying it in its natural and original form, possibly over many centuries, we came to see or discover ways in which the application of the rule needed to be constrained. There is of course no interesting sense in which the canons of inductive logic, statistics, and probability governing sample size and representativeness were selected *for*. But if you aim to accrue firmly grounded knowledge, to do science, or to make accurate predictions in elections, you better apply the relevant canons. In this non-biological sense the canons are adaptive, as are the abilities to read and write. Applying the norms of good reasoning yields firmer and more accurate knowledge and literacy makes for richer and more pleasant lives than do the alternatives. But neither excellence as a sophisticated reasoner nor literacy is interestingly fitness enhancing. Indeed, the best predictor in the modern world for low birth rate is the average level of education attained, the two having an inverse relation. The main point is that a trait, such as being a whiz at applying the sophisticated canons of inductive reasoning or being an avid reader or writer, can be rightly understood as being adaptive in the sense that possessing the relevant trait or ability contributes to knowledge, flourishing, happiness, and the like, without its being adaptive in the sense that it was selected *for* because it contributed to inclusive genetic fitness.

Now I want to claim that the same sort of point applies in the moral case. Some of the reactive emotions on Strawson's list require development, discovery and canonization over some segment of world historical time. Feelings of pride, dignity, and respect, I have suggested, fit this bill. They require development of a certain conception of a person, of norms governing behavior, do's and don'ts, oughts, institutions governing moral praise and blame, and methods for punishment of those who stray too far from the right path.

As is the case with the canons of reasoning that were discovered and developed and that can be used to constrain our original and natural inductive propensities and thus to enhance our abilities to know, so it is with morality. It is an interesting question whether we should think of the moral conceptions that are developed and the more complex attitudes that figure in living according to the moral conception as involving modifications on the original equipment or the creation of wholly new equipment. I am inclined toward the former view, but am not positive that it is the correct view. One reason I am inclined toward the former view is because I have not been convinced by the argument of a very interesting book entitled *Unnatural Emotions*, by Catherine Lutz, an anthropologist and social constructionist. Lutz argues plausibly that the Ifaluk who live on the Caroline Islands in the South Pacific have moral emotions different from us Americans. In one case, the case of the reactive attitude of *song*

which Lutz translates as justifiable moral anger, the Ifaluk do experience and express *song* somewhat differently and in response to somewhat different situations from the ways we do. The Ifaluk expect and display greater gentleness than your average American and live according to less individualistic norms than we do. So what a native New Yorker would hardly consider rudeness would be judged as a violation of norms of expected gentleness, which would lead to *song* and which might include as part of its display a refusal to eat. An angry American might inadvertently forget to eat or not feel like eating if she is morally outraged, but it is not an expected component of the expressive display of American moral anger. In any case, despite being tuned differently from American moral anger, Ifaluk moral anger is very familiar; it belongs to the same family as our anger. The second case, Lutz discusses is an emotion called *fago*. The Ifaluk use *fago* to express love, compassion, and sadness. One might think there are actually three words for three emotions here and that the Ifaluk disambiguate the homonyms by context, so that *fago* uttered in a romantic situation means love, at a funeral it might be an expression personal sadness or if uttered to the grieving spouse it might be understood as a display of compassion. Lutz interestingly argues that there is actually only one complex emotion here, one that combines feelings of love/sadness/and compassion all at once. The short explanation for why the Ifaluk link these three emotions has to do with certain historical facts about the fragility of their lives due to widely remembered bouts with deadly disease and deadly weather. When a Ifaluk looks his beloved in her eyes and declares his love for her, he simultaneously experiences feelings associated with prior losses of his own or of his beloved, he recognizes that eventually he will, in all likelihood, be lost to her or her to him, or even if they do not lose each other, their children might die young, and so on. This is all interesting, if true. But again it doesn't look to me as if we are dealing with new culturally created moral emotions here, but only with a somewhat unusual configuration of already familiar ones.

Modifying and moderating the reactive attitudes

Early in this paper I asked a set of questions:

1. Did the reactive attitudes arise as *adaptations*?
2. If so, are they still adaptive?
3. How *modifiable* are these attitudes?
4. If they are modifiable, are there any reasons to work to modify, moderate, or otherwise adjust how, when, and under what circumstances we experience or express the reactive attitudes?

I've argued that the answer to the first question is 'yes'. At least many of the reactive attitudes, although not all of them, on Strawson's list are variations on or reasonable facsimiles of Ekman's basic emotions and can thus be reasonably viewed as original and natural, and furthermore as adaptations, as dispositions that were selected *for* because they causally contributed to the sort of interpersonal commerce we engaged in when we evolved as a distinctive species. I want to emphasize now, however, something I said only in passing earlier. In saying that the reactive emotions are original and natural and adaptations to boot is not to say that they evolved *in* *Homo sapiens*. It is possible that the reactive emotions, pretty much in the form we are naturally disposed to display them, were delivered by gene sequences that belonged to and evolved in ancestral species, in earlier groups of hominids or in some non-hominid ancestor.

This possibility in itself might make one think that given that our brains and bodies as well as our social structures differ from that of our ancestors, then just being handed over some equipment these ancestors possessed might not be optimal. Just as the tinkering with the Panda's wrist bone that produced the Panda's thumb did not produce the best possible thumb for picking bamboo leaves, a direct handoff of the reactive attitudes of Neanderthals, say, may have been the best Mother Nature could do under the constrained circumstances governing the speciation event that brought *Homo sapiens* on the scene, but not have been optimal in either the original evolutionary situation nor in the changed circumstances of later, cultured environments that humans would create and inhabit.

This sort of possibility, based on knowing the twin facts that an adaptation can cease to be fitness-enhancing if and when an environment in which it was originally fitness-enhancing changes enough, *and* that Mother Nature often sacrifices even when it comes to adaptive designs, makes me hesitant to assert wholeheartedly that our original and natural reactive attitudes are still adaptive, even if we restrict the meaning of 'adaptive' to fitness-enhancing. Answering this question requires knowing a whole bunch of things about which I for one only have some hunches, hunches about which I am not particularly confident. For example, we will want to know certain things about the standard intensity, if there is such a thing, of the reactive attitudes. How strong or weak are the original settings? We will also want to know what sorts of situations standardly elicit the reactive attitudes. Saying that they are elicited by benevolence and malevolence says little until we know what sorts of things were perceived originally as benevolent or malevolent and which ones are now. We can make some plausible, educated guesses here. But much information is missing,

especially once culture is introduced, because we know different cultures conceive of benevolence and malevolence differently.

Despite lacking this information we might still be inclined to say that in current environments, especially with certain technologies at our disposal, we need to be wary of certain of the original and natural reactive attitudes. Expressions of anger in an environment filled with guns has, all else being equal, more dangerous and more deadly potential than in a world in which the standard expressions can only go as far as fists and sticks. Many people worried, and worried rightly, that being poised, even willing to fight a nuclear war that might well have resulted in species extinction was caused in part by the facelessness of the enemy. Remember that we are attuned to feel emotions of anger as well as compassion to faces, but not to large chunks of, even inhabited, land on maps. On the other side, mass communication, it is often said, gives face to suffering; we see starving children half way across the globe and are, at least sometimes, moved to help. Having said this much leads in the direction of a mixed verdict about whether in the world as we know it the reactive attitudes are well-suited for doing the job for which they were designed.

What we do know is how to answer (3) and (4). Regarding (3), the reactive attitudes are modifiable. The evidence for this claim abounds. Contemporary moral educational practices aim at and sometimes succeed in moderating what are judged to be excessively angry displays, and benevolent dispositions can be developed and enhanced although it is a variable matter how hard we try to do so. Different cultures, different moral communities work in different ways to increase or decrease guilt, and so on. The reason (4) gets a positive answer as well — namely that should we work to modify, moderate, or otherwise adjust the original reactive attitudes — has to do with the fact that there came a time every place on earth when a conception of a person and a conception of a morally good life was discovered, developed, and articulated. Every conception I am familiar with, whether the wisdom is that contained in the *Torah*, in *The Old and New Testaments*, in Confucius's *Analects*, in the *Puranas* and the *Bagahvhad Gita*, in the *Koran*, in Buddhist texts, or in secular moral theory of the sort we get from Aristotle, Mill, and Kant in the West, puts forward all sorts of wisdom and advice about how we ought to structure our cognitive-conative economies, how best to live a life, what virtues are the best expressions of our common humanity and which feelings and vices we need to be most watchful of and ready to fight off. It seems to me that each of the traditions I have mentioned, despite sometimes displaying parochial, xenophobic, sexist, and racist attitudes, do identify problems with living our lives according to our

biological natures and provide considerable wisdom, each in their own way, for being better that we are naturally prone to be, even wisdom that might help us to be good — really good.

There is a tendency to think that what is natural and original is unmodifiable. Recent work in evolutionary psychology does seem to me, at any rate, often to look too resolutely at our seamy side. It highlights certain unsavory tendencies we may in fact have in virtue of being animals, and to make it seem that because these traits are part of our nature, for example, male promiscuity, they will be very hard, if not well nigh impossible, to take control of, to modify suppress, or redirect. And Antonio Damasio has recently suggested that controlling an emotion has about as much chance of succeeding as controlling a sneeze. But again the evidence abounds that we are plastic, and that both culture and we as individuals can work with our natural equipment, as our norms require.

Conclusion

When P.F. Strawson published “Freedom and Resentment” in 1962 he was worried about an impasse he saw in moral philosophy. Moral philosophers were writing as if they needed to figure out the right answer to the metaphysical question of free will-determinism before they could judge expressions of moral approval or disapproval to be warranted. The warrant for judging people to be responsible for their actions, traits, or characters would depend on whether they could or could not control their actions, traits and characters. And both the form and rationale for our practices of expressing approval and disapproval would need to be guided by judgements about whether people could in fact choose to do or be other than they in fact did or were.

When His Holiness, the 14th Dalai Lama, writes and says that many emotions display excessive craving, excessive caring about ephemeral things, and are worth eliminating, he is also speaking from a perspective steeped in metaphysics and moral theory. There are views about the nature of the self and identity and goodness operating in every vicinity.

As for what I have called “intrinsically destructive emotions”, those that have negative valence and cause unease and disequilibrium to her who expresses or is the recipient of such an emotion, Buddhism hold out the possibility of experiencing a negative emotion without *identifying* with it, thereby removing oneself from its destructive influences. On this view, a negative emotional

experience, so long as the agent has done work or has cause not to identify with it, may not, or need not, be such that it produces the negative affect or unease I have identified as the source of intrinsic destructiveness. Assuming that the decoupling of the negative experience, and identification with it, can be pulled off in practice (the meaning of “identification” needs, of course, to be spelled out) would imply that such emotions are not intrinsically destructive simply in virtue of having negative valence.

Furthermore, and relatedly, and to make things more complicated for the analysis of destructiveness I have started to sketch, Buddhists make a distinction between afflictive and nonafflictive emotions. Afflictive emotions are not all experienced as “negative”; but their influence on one’s mental, and possibly physical, well being is still detrimental. The so-called primary mental afflictions, according to Buddhism, are ignorance, attachment, and hostility. Attachment tends to arise in response to pleasurable stimuli, and it may feel “positive”, but it actually disturbs the equilibrium of the body/mind, creating agitation, anxiety, etc. On the other hand, in Buddhism, not all emotions that feel bad are necessarily afflictive. Hostility tends to arise in response to disagreeable stimuli, especially to perceived acts of malevolence, and it, too, disturbs the equilibrium of the person. But if I have found a way not to allow hostility to afflict me, perhaps by working to diminish the hold it takes of me, or to diminish the degree to which it takes hold, or by distancing myself, or not identifying with the hostility, it may not afflict me (at least not much).

One area where western moral philosophy does, *prima facie*, seem to diverge from certain aspects of Buddhist thinking about the emotions is as follows: work like Strawson’s on the reactive attitudes, and those of others working the same terrain, despite acknowledging that these emotions can be excessive or deficient (and thus following Aristotle in thinking that there is a mean, a middle way, that we want to situate these emotions at, still believe that expressing and feeling such emotions as indignation, moral outrage, regret, remorse, guilt, compassion, gratitude, caring, concern, love, and attachment, etc., are essential to expressing our humanity, as well as to acknowledging the humanity of others. Thus work at moderating and modifying these emotions needs to proceed cautiously and *eliminating* any of them is, not to put too fine a point on it, downright wrong (as well as, practically impossible). But perhaps our picture is just that — ours — and cherished primarily for its powerful roots in our history and not because it is rooted in deep wisdom, in the best picture of normative self-construction. Maybe.

Bernard Williams makes a useful distinction between real and notional

options and about the dangers of judging across radically different life forms. I am, as I said at the start, very much interested in the Tibetan Buddhist take on destructive emotions. Following Williams's wisdom, however, I won't say any more than that I personally find the idea of eliminating emotions odd. Is the idea wrong or crazy? Of course not. And it is interesting that it comes from a non-rationalist tradition. There is more, much more to be said, but I will not say it here. Just one final thought: what I like about Tibetan Buddhism and about us is that we are engaged in the project of finding methods for "overcoming the genotype." We are animals to be sure, but unusual ones, capable of breaking out of, adjusting, and modifying what Mother Nature has tried to make us be. Of course, Mother Nature, perhaps inadvertently, has given us these gifts as well. They are precious.

Notes

1. Cleaner examples involve cases where the same gene sequence carries resistance in a type of grass to certain toxins (which is an adaptation) as well as coding for a certain shade of color in that grass. Supposing that the shade of the grass is neutral from the point of view of fitness, we can say, then that there was selection *for* toxic resistance and selection *of* color.

References

- The Dalai Lama and Howard C. Cutler M. D. (1998). *The art of happiness*. New York: Penguin Putnam.
- Damasio, Antonio. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Putnam.
- Damasio, Antonio. (1999). *The feeling of what happens: Body and emotion in the making of consciousness*. New York: Harcourt Brace.
- Darwin, Charles. (1872/1998). *The expression of the emotions in man and animals. With an introduction, afterword and commentaries by Paul Ekman*. London: Oxford.
- Ekman, Paul, and Rosenberg, Erika (Eds.) (1998) *What the face reveals*. New York: Oxford.
- Gibbard, Alan (1990). *Wise choices, apt feelings*. Cambridge: Harvard.
- Gould, S. J. (1980). *The panda's thumb*. New York: W. W. Norton.
- Griffiths, Paul E. (1997). *What emotions really are*. Chicago: Chicago
- Lutz, Catherine. (1988) *Unnatural emotions*. Chicago: University of Chicago Press.
- Neadar, Karen, (1999). Fitness and the fate of unicorns. In Valerie Gray Hardcastle (Ed.), *Where biology meets psychology: philosophical essays*. Cambridge: MIT press.

Sober, Elliot (1984). *The nature of selection*. Cambridge: MIT Press.

Sober, Elliot and David Sloan Wilson (1998). *Unto others*. Cambridge: Harvard.

Strawson, P.F. (1962) Freedom and resentment. *Proceedings of the British Academy*, vol. 48.