

Chapter Seven

Misleading Emotions

Peter Goldie

Over recent years there has been an optimistic trend in empirical and evolutionary psychology, and also more widely amongst philosophers of a broadly 'cognitivist' bent, which emphasizes the usefulness of emotion in picking up saliences in the environment, and enabling quick and effective action with little or no conscious deliberation. This optimism, I believe, deserves to be tempered with some realism. The emotions can systematically mislead us, and they can do so in ways that are systematically hard to detect and correct. This is especially the case in respect of those emotional capacities which evolved in environments that differ in important respects from the environment in which we now live.

1. Introduction

It is an orthodox view these days, and a correct one I think, that our emotional capacities and dispositions enable us to perceive certain things in our environment as salient, and to respond emotionally to those things in ways that are advantageous for us (see, for example, Elgin, this vol.; Tanesini, this vol.; de Sousa 1987; Griffiths 1997; Greenspan 2006). Roughly speaking, we are better off, as individuals and as a species, in having such capacities and dispositions. For example, when we are in a dangerous situation, we are able to perceive things as dangerous, and to respond appropriately with fear and with evasive action, doing so at a speed and with a lack of conscious deliberation that we would not be able to achieve without our capacity for fear. It may well be that, as Joseph LeDoux (1998) has argued, we have a 'high road' and a 'low road' for processing this kind of information – a more complex, slower route, and what is sometimes called a 'quick and dirty' or 'fast and frugal' route. This idea is now taken very seriously by empirical psychologists in the form of what is often called 'dual process' theory. The two processes have various names, but I will call them 'intuitive thinking' and 'deliberative thinking'.¹ They are meant to complement each other. Intuitive thinking, involving emotion and imagination, operates fast, does not involve conscious thinking and plays a vital epistemic role in a world in which energy resources are limited and speed of response is of the essence. Deliberative thinking, in contrast, is more 'cool', involves conscious deliberation and has as one of its functions operating as a check or balance on intuitive thinking.

1 See, for example, Haidt (2001); Haidt (2007); Greene et al. (2001); Gigerenzer (2000). I discuss some related issues in Goldie (2007).

But both processes are cognitive; the old idea, if there ever was such an idea, of the emotions as mere irrational or arational urges is long gone.

I willingly accept the idea that emotions and intuitive thinking can and do play this role of helping us to manage our way through the world under constraints of finite time and energy resources. But I want to add a more pessimistic note about the role of deliberative thinking as check and balance. The emotions, I will argue, can and often do *systematically* mislead us, and this gives rise to some interesting and practically challenging epistemic issues which, as will emerge, have some connections to the so-called heuristics and biases tradition in empirical psychology.² The ‘systematically’ is important here. Of course there is nothing fundamentally problematic with the occasional misleading emotion – for example, with the occasional fear felt towards something that is not really dangerous. It might get you into difficulties at the time, but it does not throw into question the role of emotion in general, any more than the occasional false belief throws into question the role of our belief-forming capacity in general. The concern that I want to raise about emotion here goes deeper than that. Furthermore, emotions not only systematically mislead us, but they also do so in ways that can be very hard to detect – and to correct. So we cannot always safely rely on epistemic checks and balances on emotion and intuitive thinking from the slower and ‘cooler’ processes of thought involved in deliberative thinking. Even if the old idea of the emotions as irrational or arational is long gone, we should avoid a recoil into the idea of the emotions as thoroughly ordered, and nicely and reliably aligned with reason. On the contrary, they can both undermine reason and disguise the fact that this is what they are doing.

2. Misleading Sallences

In general, our emotional capacities and dispositions ‘filter’ our perception of our surroundings (Wollheim 1999; Roberts 2003; Döring, this vol.), and to that extent emotional salience and emotional arousal are generated according to what those capacities and dispositions are. So, for example, a man who is disposed to find cows frightening will see them as frightening, dangerous and liable to stampede in ways that the rest of us would not when confronted with a herd of cows, and he will feel fear where we would not. Salience and arousal go together.

Salience and emotional arousal can be distorted because of irrational emotional dispositions, and I dare say that a morbid fear of cows is one such. But I am not concerned here with phobias, obsessions and the like. I am more interested in those cases where there is nothing irrational about the emotional disposition as such, but where it can still be systematically misleading. To see what I mean, let us start with David Hume, who made the point in relation to our moral sentiments. We will see that the point turns out to be quite general.

‘Morality’, Hume famously said, ‘is more properly felt than judged of’ (Hume [1739/40], 470).³ But he made it clear that not *any* feelings are relevant

² See, for example, Nisbett and Ross (1980), and Kahneman, Slovic and Tversky (1982).

³ Throughout I have modernized Hume’s spelling.

in distinguishing moral good and evil. If we are to engage in moral thought and discourse, two kinds of correction, using reason and imagination, are required. First, in judging the morality of an action we must put to one side the feelings we have which result from the contingencies of our own particular relations with the protagonists. For example, the fact that the Good Samaritan happens to be my son should not influence my moral judgement of his action, even though his kindness and generosity are more salient for me, and even though I accordingly feel more admiration, just *because* he is my son (Hume [1739/40], 472). This is not to say that I should necessarily avoid having the feelings proper to my particular relationship with my son, but rather that these feelings should be kept apart from my feelings about the morality of his action. Indeed, as Hume observed, these two kinds of feeling can even be contrary: 'The good qualities of an enemy are hurtful to us; but may still command our esteem and respect ... as when the fortifications of a city belonging to the enemy are esteemed beautiful upon account of their strength, though we could wish that they were entirely destroyed' (Hume [1739/40], 586–7). So the first thing we have to take into account and adjust for is the particularity of relationships: the same act of kindness done by a friend, an enemy or a complete stranger should be judged to have the same moral properties even though the salience of the action's moral properties, and our moral feelings in response, are different.

Secondly, our moral sentiments vary depending on the contingent proximity or remoteness of the object of our sentiments, and we need to correct for these variations in sentiment too when arriving at moral judgements of virtue and vice (Hume [1739/40], 581). What is nearer is more salient than what is farther away, and thus has a greater effect on our sentiments. Hume draws this analogy: our judgement of the moral esteem of someone distant should be corrected by reason, just as we should judge that an approaching object is not really getting any larger even though it might appear to be getting larger (Hume [1777], 227–8).

So reason has an important role here, to correct for the biases to our moral judgements that can arise from these two influences on our moral sentiments. (In this sense one can see him, perhaps rather controversially, as a kind of proto-dual-process theorist; more of this later.) If our moral judgements and discourse were slavishly to follow our sentiments, as Hume says, 'it is impossible we could ever converse together on reasonable terms', and '[i]n order, therefore, to prevent these continual *contradictions*, and arrive at a more *stable* judgement of things, we fix on some *steady* and *general point of view*; and always, in our thought, place ourselves in them, whatever may be our present situation' (Hume [1739/40], 581–2). Hume made the point many times over, both in the *Treatise* and in the *Enquiry*, that this requirement, to take a steady and general point of view, is essential to our moral discourse.⁴

4 'It is impossible men could ever agree in their sentiments and judgements, unless they chose some common point of view' (Hume [1739/40], 591). 'The intercourse of sentiments ... makes us form some general inalterable standard' (Hume [1739/40], 603; Hume [1777], 229). 'General language, therefore, being formed for general use, must be moulded on some general view' (Hume [1777], 228). 'The notion of morals implies some sentiment common to all mankind, which recommends the same object to general approbation, and makes every man, or most men, agree in the same opinion or decision concerning it' (Hume [1777], 272).

But it is at this point that Hume reveals his pessimism about the corrective power of reason. We might manage to make moral ‘pronouncements’ based on our reasoned adjustments to our sentiments, in line with the general requirement to take a steady and general point of view, but nevertheless the heart seldom follows the head: ‘I do not feel the same lively pleasure from the virtues of a person who lived in Greece two thousand years ago that I feel from the virtues of a familiar friend and acquaintance’ (Hume [1739/40], 581). Whilst reason might enable me to judge that my friend and the distant Greek are equally virtuous, my close relationship with my friend, and his proximity, still have their distorting influence on my passion: the general principles for correction by reason ‘are not altogether efficacious, nor do our passions often correspond entirely to the present theory’ (Hume [1739/40], 583). If our passions do not entirely fall into line with correction by reason, then, one might think, it is likely that our motivations will also follow our heart and not our head – that they will follow the direction of our sentiments and not that of our judgements or discourse. I will return to this important point shortly.

The question that I now want to address is whether Hume’s idea of how sentiments or passions can bias judgement generalizes beyond the moral domain. Do the same biases arise elsewhere too? According to the work of the psychologists Richard Nisbett and Lee Ross, and of Daniel Kahneman and Amos Tversky, founders of the heuristics and biases tradition, indeed they do. We can immediately see the connections between their work and Hume’s moral philosophy. Let us consider Nisbett and Ross, who, in Chapter 3 of their *Human Inference: Strategies and Shortcomings of Social Judgment* (Nisbett and Ross 1980), discuss how emotions can have misleading or distorting effects on our inferences according to what they call the ‘vividness criterion’. Information is vivid, they say, when it is ‘likely to attract and hold our attention and to excite the imagination to the extent that it is (a) emotionally interesting, (b) concrete and imagery-provoking, and (c) proximate in a sensory, temporal or spatial way’ (Nisbett and Ross 1980, 45).

Regarding (a), emotional interest, events tend of course to have more emotional interest if we are ourselves directly involved. And, as Nisbett and Ross add, ‘events also are more interesting when they happen to people we know than when they happen to people we do not know, and they are more interesting when they happen to people about whom we have strong feelings than when they happen to people about whom we have neutral feelings’ (Nisbett and Ross 1980, 46). Regarding (b), a concrete and imagery-provoking event will be one which, roughly, has more emotional content in virtue of the way what happened is grasped; the more emotional detail there is available, the more salience and impact it will have. For example, a short report in the newspapers of a flood in the Philippines with five thousand dead will have less emotional detail than a report on television with close-up footage of the devastation and of the drowned bodies. And discussing the proximity of information, (c), Nisbett and Ross give this example: ‘The news that a bank in one’s neighbourhood has been robbed just an hour ago is more vivid than the news that a bank on the other side of town was robbed last week’ (Nisbett and Ross 1980, 49). These three elements of the vividness criterion are typically found together, even if they are conceptually distinct (Nisbett and Ross 1980, 45).

Nisbett and Ross go on to argue that the vividness of information affects inferences. (The epigraph to the chapter is a quotation from Bertrand Russell: 'Popular induction depends upon the emotional interest of the instances, not upon their number'.) When information is what they call 'pallid', as dry statistical evidence is, it tends to be given less inferential weight than the evidence of one's own eyes or evidence that is otherwise more salient. The robbery in one's own neighbourhood an hour ago, being more vivid than the robbery last week on the other side of town, 'is likely to have a greater impact on one's views of the seriousness of the crime problems in one's city or the need for stiffer prison sentences for bank robbers' (Nisbett and Ross 1980, 50). And the vivid television footage of the flood in the Philippines is likely to make one more concerned about the inadequacy of the flood defence systems than is the pale newspaper report.

A direct line-by-line comparison of Hume's account with that of Nisbett and Ross would not be fruitful without a lengthy prior examination of their philosophies of mind and of action. Nevertheless, it seems clear that what their accounts have in common is the idea that our judgements and inferences about empirical facts can be systematically biased by emotion in cases where emotions pick up on certain saliences that are not relevant for the judgement or inference. The question then is how these biases affect motivation.

Sometimes, of course, vividness of information can have a beneficial effect on motivation, getting one to do something that one ought to have done anyway, but never got around to. Nisbett and Ross give a nice example of how, in 1974, the much-reported and much-discussed mastectomies performed on Mrs Ford and Mrs Rockefeller led to a mass of visits for cancer checkups, whereas in the past widely disseminated statistics about the risks of breast cancer had produced nothing like such a significant effect on behaviour. But saliences do often mislead one into error, leading one to attach too much weight to certain considerations and thus leading one to act wrongly—ethically or prudentially.

Towards the end of their chapter on misleading saliences and the vividness criterion, Nisbett and Ross accept, as I do on my own behalf, that they have chosen to emphasize the 'serious inferential and behavioural costs' associated with emotional saliences, the reason for this pessimism being, simply, that '[t]he vividness of information is correlated only modestly, at best, with its evidential value' (Nisbett and Ross 1980, 60). But they then go on to make some speculative remarks about the reasons why we are subject to making these errors in reasoning and behaviour as a result of emotional saliences. This will help to lead us towards the next thing that I want to discuss – how misleading emotional saliences arise through environmental mismatch. This is what they say:

During all but the most recent moments of our evolutionary history, dangers and opportunities have been relatively concrete and vivid. ... Now, however, our world has come to have pressing dangers which are complex and abstract matters ... , dangers best described by abstract and often statistical information. (Nisbett and Ross 1980, 60)

The suggestion, then, is that what used to be advantageous is now not so, because the environment has changed in respect of its increasing complexity, and in respect

of our increased knowledge of its complexity. We will see that the environment has changed not only in this way, but in others too.

3. Environmental Mismatch

Let us begin with the notion, recently developed by Gerd Gigerenzer and his collaborators, of ‘bounded rationality’, and see how it relates to the heuristics and biases tradition of Nisbett and Ross and Kahneman and Tversky. The central idea of bounded rationality, emerging from the discipline of evolutionary psychology, is that humans, under the constraints of time and limited information, do not typically reason using the accepted standards of decision theory. We act on the strength of a collection of rules and heuristics embodied in what they call an ‘adaptive toolbox’, which has the following features:

First, it [is] a collection of rules or heuristics rather than ... a general-purpose decision-making algorithm Second, these heuristics are fast, frugal, and computationally cheap rather than consistent, coherent, and general. Third, these heuristics are adapted to particular environments, past or present, physical or social. ... Fourth, the bundle of heuristics in the adaptive toolbox is orchestrated by some mechanism reflecting the importance of conflicting motivations and goals. (Gigerenzer and Selten 2001, 9)

Emotions, they say, can themselves be heuristics – let us call them ‘emotion-based heuristics’. Gigerenzer’s example is disgust, which operates only within the domain of things we choose to ingest, and which serves the adaptive function of preventing poisoning (Gigerenzer 2001, 42). If the waiter in the restaurant presents us with a green steak on the plate, we will feel disgust towards it and refuse to eat it, even if we are assured that it is harmless colouring.⁵

We can see that there is agreement between the recent work of Gigerenzer and colleagues on bounded rationality, and that of Nisbett and Ross and Kahneman and Tversky in the heuristics and biases tradition, in this respect: they agree that our emotion-based heuristics have a very important role in our intuitive responses to the world around us, and as a result in the ways we act. However, whereas the earlier researchers put the emphasis on the errors and biases that can arise, Gigerenzer and colleagues say that ‘bounded rationality is not an inferior form of rationality’ (Gigerenzer and Selten 2001, 6). Peter Todd, writing in the same volume, registers the disagreement by commenting that ‘the basic message of [the heuristics and biases] research program ... is that humans use heuristics at their peril, more often than not making errors of judgment and inaccurate decisions. ... In contrast, the vision of ecological rationality emphasizes that humans use specific simple heuristics because they *enable* adaptive behaviour, by exploiting the structure of information in natural decision environments. Simplicity is a virtue, rather than a curse’ (Todd 2001, 52–3). So the rationality wars, as they are sometimes called, are not over the

⁵ See also Todd and Gigerenzer (2000, 740 and 741) where the precise role of emotion is discussed as one of the ‘challenges that remain’. Adam Morton (2000) discusses the possibility that a heuristic can also play the role of ‘orchestration’.

role of emotion in our reasoning, nor are they over the emergence of our emotions from our evolutionary past. Rather, the rationality wars are, at least in part, over whether or not heuristics, including emotion-based heuristics, in general serve our interests *in the environment in which we now live*: in other words, whether we should be pessimistic or optimistic about the epistemic role of heuristics, and thus, we can add, of emotion-based heuristics in particular.⁶

The answer to this question will depend, in part, on whether or not there is a match between, on the one hand, the actual environment in which we now live and, on the other hand, the related domain-specific emotion-based heuristic. ‘Heuristics that are matched to particular environments allow agents to be *ecologically rational*, making adaptive decisions that combine accuracy with speed and frugality. The degree to which a match exists ... determines how accurate a heuristic is’ (Gigerenzer 2001, 46). I will argue, in respect of at least three types of emotion, that there is a *systematic* environmental mismatch, that this mismatch *systematically* leads to wrong intuitive thinking, and thus to wrong motives and wrong actions, and furthermore, that this mismatch is *systematically* not easy to detect or correct through reason, through deliberative thinking. To repeat what I said at the outset, the concern I want to raise is not the bare possibility of what are called ‘false positives’. These can be no bad thing: better to jump at a stick on the path, fearfully taking it to be a snake, than not to jump at a snake on the path, blithely taking it to be a stick. The concern goes much deeper than false positives.

Let us start with male aggression. The anthropologist Daniel Fessler tells of a case from his fieldwork in south-western Sumatra which illustrates well what I have in mind. Late one night a man, Rustram, and his girlfriend hired a minibus to take them home. In a senseless dispute over the fare (the discrepancy was tiny by any standard), a fight began and the driver and his friend stabbed Rustram fourteen times and left him for dead by the roadside with his girlfriend. Discussing what happened, some of the villagers blamed it on possession by the Devil, but another wisely said it was because of *malu*: ‘No one wants to be *malu* in front of a girl’ (Fessler 2001, 194). Literature, history and empirical psychology are replete with such examples, and we have all witnessed or been involved in incidents of men getting wildly angry and out of control; something like it happens most nights in most big cities, so often over some trivial matter, such as who was first in the queue at the bar.

Malu is an emotion which is something like a simpler version of shame, and its converse, *bangga*, is something like a simpler version of pride – Fessler calls them Protoshame and Protopride. They are based on what he calls a three-point logic, in which the other ‘is viewed not as a target for intersubjectivity, but merely as a

6 J.-L. Bermúdez (2000) discusses an important issue, which I will not directly address: what the status of bounded rationality is in normative epistemology, as contrasted to descriptive epistemology; in other words, whether or not Gigerenzer and other proponents of bounded rationality are claiming that this is the way we *ought* to reason, as well as claiming that this is the way we, in fact, do reason. For an irenic discussion of the rationality wars see Samuels, Stich and Bishop (2002), and Samuels and Stich (2003).

feature of the social world' (Fessler 2001, 197), and can thus be emotions which are homologous with emotions in other animals.⁷

Although Protoshame and Protopride (unusually, Fessler remarks, English does not have a term for them) are thoroughly dysfunctional for human beings in today's environment, they were not always so. Fessler speculates on their adaptiveness:

Because Protoshame, an aversive emotion, is elicited by subordination, while Protopride, a rewarding emotion, is elicited by dominance, individuals capable of experiencing these emotions would have been motivated to seek out higher rank ... [and hence] would have greater reproductive success than those who lacked these traits. (Fessler 2001, 198)

We need not be committed to this particular explanation of why Protopride or Protoshame are selectionally advantageous, of why they are 'adaptations' or products of natural selection which evolved for adaptive reasons over the species' evolutionary past (Barkow, Cosmides and Tooby 1992). All we need is the idea that they play a role in the emotion-based heuristic concerned with male aggression towards other males, and that in an ancestral environment they were adaptive – for some reason or other. But today, of course, we have an environmental mismatch. Not only are violent disputes between males over trivial matters often ethically problematic, with the dominant man taking advantage of his physical superiority. This kind of aggression also now characteristically fails to serve the purpose that it did: disputes get out of control with no one the winner in the end, often also against the best interests not only of the individuals concerned but also of the wider community. At the time of the dispute the protagonists will do everything in their power to win the day, sacrificing much else of what they value when they are not in the grip of the emotion: their health, their liberty, their dignity and much else besides.⁸ Leaving all this to one side, there is also the terrible damage that can be done to someone's psyche through sublimation of his aggression against other men into some other activity.⁹

The second kind of domain-specific emotion-based heuristic which has an environmental mismatch is fear and mistrust of strangers – xenophobia. The xenophobe tends immediately and unreflectively to react adversely to those who are 'not like us', and to treat them with suspicion or even worse. The 'fast and frugal' nature of such responses has some quite robust support from the well-known Harvard implicit association tests for race, in which we are found to respond much more quickly to terms such as 'good' and 'nice' when juxtaposed with white faces than we do when they are juxtaposed with black ones.¹⁰

7 The three-point logic for Protoshame is (1) the other is assessed as more important than the self; (2) the self must interact with the other in which (1) will be salient; and (3) as a result the self feels Protoshame or *malu* (Fessler 2001, 195–6).

8 One might argue that what we have in such cases is not irrationality but simply radically shifting preferences (see Elster 1999b), but this is surely rescuing rationality at the cost of common sense. See also Fessler (2001, 208–9).

9 These emotions can also be economically damaging, with males competing pointlessly to have the 'best' positional good such as the biggest house or the fastest car.

10 See <https://implicit.harvard.edu/implicit/>.

Jim Hopkins has argued that the tendency to what he calls 'ingroup cohesion and outgroup conflict' appears in individual motivations as 'two sides of the same evolutionary coin' – once adaptive, now dysfunctional, and, however much 'we moderns think ourselves guarded against such modes of thought, the roots remain active' (Hopkins 2004, 225). The evidence for this is all around us, in conflicts as far apart as local gang wars in South London and tribal disputes in remote parts of Africa.

Finally, there is male sexual jealousy, which, like male aggression, is supposed to be pan-cultural and shared with other animals. Let us assume, as a number of evolutionary psychologists have argued, that the capacity for male sexual jealousy is selectionally advantageous (see for example Wilson and Daley 1992); again the details of the explanation do not concern us here. This would then explain why we now have this domain-specific emotion-based heuristic, and it is, indeed, a very clear example of the kind of fast and frugal thinking that Gigerenzer and colleagues have in mind.¹¹ It involves an amazing capacity to pick up the saliences of a possible or imagined transgression: the slightest sign, 'trifles light as air', can be enough. The mismatch that is at issue here is, as with male aggression, in part prudential but most of all it is an ethical one: we now know that male sexual jealousy, with its inbuilt capacity to get violently out of control, is wrong (see Taylor 1988; for a nuance on Taylor see Goldie 2000). In Patrick Marber's play, *Closer*, the doctor Larry blames his behaviour on the fact that he is just a 'caveman'; but this is not even an excuse, let alone a justification.

With this discussion in mind, we can now return to our earlier discussion of morality, and see indications here too of environmental mismatch in the influence on moral sentiments of the particularity of our relationships and of the proximity of the object of our moral attitudes. Examining these connections in the light of recent work in moral psychology on dual process thinking will reveal why environmental mismatch in general, whether in the domain of morality or elsewhere, gives rise to specifically *epistemic* concerns and difficulties.¹²

In a number of important papers, Paul Slovic, writing in the heuristics and biases tradition, has argued that a principal reason why we fail to act when we become aware of genocide is because of failure to be properly *affected*: 'the statistics of mass murder or genocide, no matter how large the numbers, fail to convey the true meaning of such atrocities. The reported numbers of deaths represent dry statistics ... that fail to spark emotion or feeling and thus fail to motivate action' (Slovic 2007, 79). Slovic's explanation of this is highly persuasive. He argues that what he calls 'System 1 thinking', or what I call 'intuitive thinking', is significantly affected by the vividness criterion. Intuitive thinking not only involves emotion but also imagination, in particular perceptual imagination, which in turn affects emotion, motivation and action so that, in accordance with the vividness criterion, the amount

¹¹ Gigerenzer makes it clear that not all heuristics are innate, so new tools can be added to the toolbox through learning. Nevertheless, '[for] some important adaptive tasks ... there would be strong selective advantages in coming into the world with at least some heuristics already wired into the nervous system' (Todd and Gigerenzer 2000, 768).

¹² I am grateful to the editors for pressing me to say more about what follows.

by which one is affected emotionally by the suffering of others typically does not increase proportionately with the numbers of people involved. 'People are much more willing to aid identified individuals than unidentified or statistical victims', as Slovic (2007, 88) says.

Slovic's explanation is that there is an environmental mismatch (although he does not use the term): 'System 1 [intuitive] thinking evolved to protect individuals and their small family and community groups from present, visible, immediate dangers. This affective system did not evolve to help us respond to distant, mass murder' (Slovic 2007, 84). Slovic adduces some surprising evidence to support the claim that we have a diminishing sensitivity to human suffering – what he calls 'psychophysical numbing' – and that this begins to have its effects at a very early point on the scale of increasing numbers. In one study, participants contributed much more to a \$300,000 fund when that fund was said to be for the treatment of just one child than when it was said to be for the treatment of eight children (Slovic 2007). This result emerged even though the members of the group of eight children were individually identified; even here more distress and compassion is felt towards the identified single child. 'Left to its own devices,' says Slovic, 'System 1 [intuitive thinking] will likely favor individual victims and sensational stories that are closer to home and easier to imagine. It will be distracted by images that produce strong, though erroneous, feelings, like percentages as opposed to actual numbers' (Slovic 2007, 91). Thus our intuitive thinking in the moral domain, and our moral motivations and actions, seem to ignore, or somehow not to be properly influenced by, what should be a plain moral truth: individual lives matter equally, whether that life is considered by itself or as one amongst a million, and whether that life is vividly salient or not.

4. Can we Correct for our Misleading Emotions?

Slovic says, following Kahneman (2003), that 'one of the important functions of System 2 [deliberative thinking] is to monitor the quality of mental operations and overt behaviors produced by System 1' (Slovic 2007, 91). The point is really a general one, extending beyond the moral domain. If we assume that humans do have, broadly speaking, dual processes of thinking, intuitive and deliberative, then Slovic is surely right that at least one of the roles of deliberative thinking is a 'monitoring' one – to act as a kind of epistemic check and balance on fast and frugal intuitive thinking. The question now arises, in relation to the systematically misleading emotions that I have been considering – male aggressiveness, xenophobia and male sexual jealousy, as well as morality – whether deliberative thinking is up to the job.

In the domain of morality, Slovic was doubtful of the motivating powers of deliberative thinking. Whatever we might judge, our actions seem to follow our intuitions. We are, as Slovic puts it, in the 'twilight between knowing and not knowing' (Slovic 2007, 82, citing Power 2003, 505). The parallel with Hume's ideas are quite striking. Hume too, as we have seen, thought that the role of reason was in part to correct for the misleading influences of our sentiments (and this was why I chose to describe him as a proto dual-process theorist). Yet he, like Slovic, was pessimistic

about reason's corrective powers, and about whether ultimately motivation and action will follow the heart or the head. As Hume put it with characteristic wit, 'though the heart does not always take part with those general notions, or regulate its love and hatred by them, yet are they sufficient for discourse, and serve all our purposes in company, in the pulpit, on the theatre, and in the schools' (Hume [1739/40], 603).

The epistemic task of unemotional deliberative thinking, to monitor and correct fast and frugal emotional-based intuitive thinking, has a place wherever the two kinds of thinking potentially clash. The task is not only for deliberative thinking to be able to sort out the good reasons from the bad, the right perspectives from the biases, and so on. It seems that, in the moral cases which concerned Hume and Slovic, it is capable of this – at least for the purposes of moral discourse. The task is also to turn these good reasons into good *motivating* reasons; and it is in just this sense, it seems, that deliberative thinking, at least in the moral domain, fails. We are indeed in the twilight between knowing and not knowing when we say, as we so often do, how terrible is the genocide reported on the latest news, and then press on to worry only about the single murder in our own home town. Categorizing the failure of deliberative thinking in this second task as a kind of 'knowing-but-not-knowing' helps one to see both that the failure is an epistemic one, and also to see that there is an important connection between the failure and weakness of the will. One might put it like this: if we *really* knew how terrible the genocide was, then we would try to do something about it.

In respect of the three other domains that I have been considering, sexual jealousy, aggression and xenophobia, the epistemic difficulties take different forms. To correct for a misleading emotion, we must first recognize it as such – that is, we must recognize that we are experiencing an emotion, and we must recognize that we are being misled by it. This can be particularly worrying, because of the propensity of misleading emotion to 'mask' its misleadingness. There are at least two ways in which it can do this.¹³

The first way in which emotions and intuitive thinking can mask their misleadingness is manifested in the emotion-based heuristic of sexual jealousy. The emotion *skews the epistemic landscape*. Trifles light as air *seem* like irrefutable evidence of transgression. And evidence which we might otherwise, through cool and calm deliberative thinking, take to count *against* our emotion we now ignore, or even take to be confirmatory of our suspicions. Leontes in Shakespeare's *A Winter's Tale* not only completely ignored what he knew to be the indisputable evidence of the oracle of Apollo (knowing-but-not-knowing again); he also saw the protests of his servant Camillo at the unfoundedness of his jealousy as confirming evidence of its justification: the servant whom he used to trust completely must now be a 'false villain', working against him and on the side of his supposed rival. (In this respect, and in others too, the epistemology of sexual jealousy is similar to that of paranoia.) One alarming thing that this example brings out is the doubtful epistemic benefit of talking to a friend, or getting advice from someone wise, in order that what they tell you should serve as a corrective mechanism. For this is

13 I will not consider here Jon Elster's (1999a) fascinating discussion of the alchemies of emotion.

just what Leontes did, and yet he turned against Camillo and ignored the oracle of Apollo. Whilst the case of Leontes may be an extreme one, I think the phenomenon is quite widespread. We so easily find the well-intentioned corrective remarks of our friends as *parti pris* or as not really properly informed (Goldie 2004).

Xenophobia too skews the epistemic landscape. The xenophobe, who believes that much of his country's troubles are due to immigrants, is faced with reliable statistics showing that this is far from the case, and indeed that immigrants have done much good for the economy. But, with his xenophobia already in place, he ignores the reliable data, and instead latches on to the vividness of the stories in his *Daily Mail* of the latest crime by blacks against property or person. Such newspapers both thrive on, and exacerbate, our only partially recognized fears. We might suggest that the xenophobe buy a newspaper which we consider to be more reliable, but why should he? After all, so far as he can tell, it is only the *Daily Mail* that has the courage to say how bad things really are, and our telling him otherwise just shows how little we know.

So the first epistemic difficulty that we face in trying to correct for our misleading emotions is that our emotionally-grounded fast and frugal intuitive thinking can systematically skew the epistemic landscape, and that we can systematically fail to recognize this. There is no contrary 'check' from cool and calm deliberative thinking because reason has *already* been undermined by emotion without our conscious awareness: the pale nonemotional evidence is *already* discounted and the vivid emotional evidence is *already* given too much weight. As Jonathan Haidt puts it, 'the reasoning process is more like a lawyer defending a client than a judge or scientist seeking truth' (Haidt 2001, 820).

The second epistemic difficulty is different. Here we might well recognize the emotion as such, and even that we are being misled by it. But now what has happened is that our preferences are skewed in the heat of the moment – the emotion has *skewed the preferential landscape*. Even if we hear the voice of cool and calm deliberative thinking telling us to stop this silliness, we are still often motivationally overwhelmed by the passion. This might sound like the old idea of reason versus the passions, but this is not so. The idea is not that passion necessarily lacks rational authority altogether; the idea rather is that the motivating power of passion often exceeds its rational authority when we are in the grip of an emotion, and the emotion itself serves to mask this fact.

Even knowing that we are in the grip of an emotion, we can still often fail adequately to recognize the error of our ways, with all the focus of our mind on the salience of the object of our passion; only too late, after the passion is spent, do we realize our mistake. This kind of misleadingness is manifested in male aggressiveness. A nice example is from F. Scott Fitzgerald's *Tender is the Night*. Dick Diver, after a night of drinking, starts a fight with a taxi driver, is arrested and then, in the police station, just as he is about to get away with a small fine, he acts stupidly, with aggression. The consequences, as he would have foreseen without the blinding passion, are terrible:

The captain stood up.

‘Écoute!’ he cried portentously. ‘Vous êtes saouïl. Vous avez battu le chauffeur. Comme ci, comme ça.’ He struck the air excitedly with right hand and left, ‘C’est bon que je vous donne la liberté. Payez ce qu’il a dit – cento lire. Va au Quirinal.’

Raging with humiliation, Dick stared back at him.

‘All right.’ He turned blindly to the door – before him, leering and nodding, was the man who had brought him to the police station. ‘I’ll go home,’ he shouted, ‘but first I’ll fix this baby.’

He walked past the staring carabinieri and up to the grinning face, hit it with a smashing left beside the jaw. The man dropped to the floor.

For a moment he stood over him in savage triumph – but even as a first pang of doubt shot through him the world reeled; he was clubbed down, and fists and boots beat on him in a savage tattoo. He felt his nose break like a shingle and his eyes jerk as if they had snapped back on a rubber band into his head. A rib splintered under a stamping heel. Momentarily he lost consciousness, regained it as he was raised to a sitting position and his wrists jerked together with handcuffs. He struggled automatically. The plainclothes lieutenant whom he had knocked down, stood dabbing his jaw with a handkerchief and looking into it for blood; he came over to Dick, poised himself, drew back his arm and smashed him to the floor.¹⁴

As a practical corrective to these epistemic difficulties, it is not sufficient simply to think about, and perhaps to ‘mouth’, the error of our ways. What matters is the excessive motivating power of the preferences, and it is not enough to counteract this just to recognize that their power exceeds their authority. This is nicely evidenced by Aristotle’s example of the emotionally weak-willed person: whilst indulging himself in food and drink, he recites the verses of Empedocles, which are all about the mistakes of excessive indulgence in food and drink. Again we are in the land of knowing-but-not-knowing.¹⁵

Another practical epistemic corrective to the skewing of the preferential landscape might be consciously to try to turn one’s attention away from the salience of the object of temptation. But this has obvious and familiar difficulties: one has to focus on the very thing that one is set on ignoring, such as, in Dick Diver’s case, the grinning, leering, nodding face of the man who had brought him to the police station.

A third epistemic corrective that one might deploy in deliberative thinking, which often has more promise of success, is not to think to oneself of the error of one’s ways (the first corrective), nor to try to reduce the salience of the temptation (the second corrective), but to increase the salience of the paler information, and thus to increase its emotional import and motivating power. In effect, this involves recruiting

14 F. Scott Fitzgerald’s *Tender is the Night*, chapter XXII of book 2; first published Charles Scribner’s Sons 1934, now in various editions.

15 See Aristotle’s *Nicomachean Ethics* Book VII, and for discussion Broadie (1991). Also see Lovibond (2002) for a more Platonic conception.

emotion, imagination and vividness on to the side of deliberative thinking. After all, there is no reason why these powerful battalions should only be deployable on the side of non-conscious intuitive thinking. Thus, for example, one might intentionally try hard vividly to imagine the bad consequences of one's action before doing it – for example imagining being seen in the act by the person you are betraying, or imagining hearing the shopkeeper's shout of 'Stop thief!' as you walk away from the scene of your petty crime. In this way the countervailing considerations become salient too, in spite of their distance, and this can then have the effect of balancing out the saliences somewhat; emotion is deployed on both sides, not only on the side of intuitive thinking. In Dick Diver's case, he might have tried vividly to imagine what would happen to him if he hit that man in that police station, and this would have given rise to fear at what he imagined, to set against the vivid fury and Protoshame that he felt at the humiliating remarks of the police captain. However, this of course required foresight and, as we have just seen, Dick realized only too late the implications of what he had done – the damage was already done – the emotions has already done their misleading work.

And this, of course, points to a fourth and very familiar corrective in which deliberative thinking can play a role: stop and think; count to ten; bite your tongue; take a deep breath; sleep on it. All these familiar everyday admonitions are implicit acknowledgements of the power of emotion, of quick and dirty intuitive thinking, to mislead us to over-hasty action.

There is nothing in principle against any of these correctives, and some or all of them surely can be effective on occasion. But there remains the central concern that emotion-based intuitive thinking does its dirty work, so to speak, before deliberative thinking comes on the scene, so that when deliberative thinking does arrive, the epistemic landscape and the preferential landscape have *already* been skewed and the dirty work *already* covered up. To adapt Haidt's metaphor, it is as if the defence lawyer, before he takes on the case, already believes that his obviously guilty client is innocent.

5. Conclusion

Over the last few years there has been this optimistic trend in emotion research, exemplified here in the work of Gigerenzer and his colleagues, and also more widely amongst philosophers of a 'cognitivist' bent, which emphasizes the usefulness of emotion in picking up saliences in the environment, and enabling quick and effective action with little or no conscious deliberation. This optimism, I believe, deserves to be tempered with some healthy realism. So too does optimism about the reliable power of deliberative thinking to correct any distortions that might arise through the influence of emotion. So when Peter Todd, writing about fast and frugal heuristics, says that 'humans are uniquely able to set aside such mental shortcuts and engage in extensive cogitation, calculation, and planning' (Todd 2001, 54), we might agree with him, so far as it goes. But I have been trying to make the realistic – or pessimistic – point that if Todd's remark, and others like it, is taken to suggest that 'setting aside' our heuristics to engage in cogitation is something we can *reliably and effectively*

do, then this optimism is open to doubt when the emotion-based heuristic has an environmental mismatch. Emotions such as these can systematically mislead us, and they can do so in ways that are systematically hard to detect, and systematically hard to correct for.

References

- Barkow, Jerome H.; Cosmides, Leda; Tooby, John (eds). 1992. *The Adapted Mind. Evolution, Psychology, and the Generation of Culture*. Oxford: Oxford University Press.
- Bermúdez, José-Luis. 2000. 'Rationality, Logic, and Fast and Frugal Heuristics. Review of Gigerenzer et al.' *Behavioral and Brain Sciences* 23: 744–5.
- Broadie, Sarah. 1991. *Ethics with Aristotle*. Oxford: Oxford University Press.
- de Sousa, Ronald. 1987. *The Rationality of Emotion*. Cambridge, Mass.: MIT Press.
- Elster, Jon. 1999a. *Alchemies of the Mind. Rationality and the Emotions*. Cambridge: Cambridge University Press.
- Elster, Jon. 1999b. *Strong Feelings. Emotion, Addiction, and Human Behavior*. Cambridge, Mass.: MIT Press.
- Fessler, Daniel. 2001. 'Emotions and Cost–Benefit Assessment. The Role of Shame and Self-Esteem in Risk Taking'. In: Gigerenzer, Gerd; Selten, Reinhard (eds). *Bounded Rationality. The Adaptive Toolbox*. Cambridge, Mass.: MIT Press. 191–214.
- Gigerenzer, Gerd. 2000. *Adaptive Thinking*. Oxford: Oxford University Press.
- Gigerenzer, Gerd. 2001. 'The Adaptive Toolbox'. In: Gigerenzer, Gerd; Selten, Reinhard (eds). *Bounded Rationality. The Adaptive Toolbox*. Cambridge, Mass.: MIT Press. 37–50.
- Gigerenzer, Gerd; Selten, Reinhard. 2001. 'Rethinking Rationality'. In: Gigerenzer, Gerd; Selten, Reinhard (eds). *Bounded Rationality. The Adaptive Toolbox*. Cambridge, Mass.: MIT Press. 1–12.
- Goldie, Peter. 2000. *The Emotions. A Philosophical Exploration*. Oxford: Clarendon Press.
- Goldie, Peter. 2004. 'Emotion, Reason, and Virtue'. In: Evans, Dylan; Cruse, Pierre (eds). *Emotion, Evolution, and Rationality*. Oxford: Oxford University Press. 249–67.
- Goldie, Peter. 2007. 'Seeing What Is the Kind Thing to Do. Perception and Emotion in Morality'. *Dialectica* 61: 347–61.
- Greene, Joshua D.; Sommerville, R. Brian; Nystrom, Leigh E.; Darley, John M.; Cohen, Jonathan D. 2001. 'An fMRI Investigation of Emotional Engagement in Moral Judgement'. *Science* 293: 2105–8.
- Greenspan, Patricia S. 2006. 'Practical Reasoning and Emotion'. In: Mele, Alfred R.; Rawling, Piers (eds). *The Oxford Handbook of Rationality*. New York: Oxford University Press. 206–21.
- Griffiths, Paul E. 1997. *What Emotions Really Are. The Problem of Psychological Categories*. Chicago: University of Chicago Press.

- Haidt, Jonathan. 2001. 'The Emotional Dog and Its Rational Tail. A Social Intuitionist Approach to Moral Judgement'. *Psychological Review* 108: 814–34.
- Haidt, Jonathan. 2007. 'The New Synthesis in Moral Psychology'. *Science* 316: 998–1002.
- Hopkins, Jim. 2004. 'Conscience and Conflict. Darwin, Freud, and the Origins of Human Aggression'. In: Evans, Dylan; Cruse, Pierre (eds). *Emotion, Evolution, and Rationality*. Oxford: Oxford University Press. 225–48.
- Hume, David. [1739/40] 1978. *A Treatise of Human Nature*. Ed. Lewis A. Selby-Bigge and Peter H. Nidditch. Oxford: Oxford University Press.
- Hume, David. [1777] 1975. *An Enquiry Concerning the Principles of Morals*. Ed. Lewis A. Selby-Bigge. Oxford: Oxford University Press.
- Kahneman, Daniel. 2003. 'A Perspective on Judgement and Choice. Mapping Bounded Rationality'. *American Psychologist* 58: 697–720.
- Kahneman, Daniel; Slovic, Paul; Tversky, Amos. 1982. *Judgment under Uncertainty. Heuristics and Biases*. Cambridge: Cambridge University Press.
- LeDoux, Joseph E. 1998. *The Emotional Brain. The Mysterious Underpinnings of Emotional Life*. London: Weidenfeld and Nicolson.
- Lovibond, Sarah. 2002. *Ethical Formation*. Cambridge, Mass.: Harvard University Press.
- Morton, Adam. 2000. 'Heuristics all the Way up?' *Behavioral and Brain Sciences* 23: 758–9.
- Nisbett, Richard E.; Ross, Lee D. 1980. *Human Inference. Strategies and Shortcomings of Social Judgment*. Englewood Cliffs: Prentice-Hall.
- Power, Samantha. 2003. *A Problem from Hell. America and the Age of Genocide*. New York: Harper Perennial.
- Roberts, Robert C. 2003. *Emotions. An Essay in Aid of Moral Psychology*. Cambridge: Cambridge University Press.
- Samuels, Richard; Stich, Stephen. 2003. 'Rationality and Psychology'. In: Mele, Alfred R. (ed.). *The Oxford Handbook of Rationality*. New York: Oxford University Press. 279–300.
- Samuels, Richard; Stich, Stephen; Bishop, Michael. 2002. 'Ending the Rationality Wars. How to Make Disputes about Human Rationality Disappear'. In: Elio, Renee (ed.). *Common Sense, Reasoning and Rationality*. New York: Oxford University Press. 236–68.
- Slovic, Paul. 2007. "'If I Look at the Mass I Will Never Act". Psychic Numbing and Genocide'. *Judgment and Decision Making* 2: 79–95. <http://journal.sjdm.org/jdm7303a.pdf>.
- Taylor, Gabrielle. 1998. 'Envy and Jealousy. Emotions and Vices'. *Midwest Studies in Philosophy* 13: 233–49.
- Todd, Peter M. 2001. 'Fast and Frugal Heuristics for Environmentally Bounded Minds'. In: Gigerenzer, Gerd; Selten, Reinhard (eds). *Bounded Rationality. The Adaptive Toolbox*. Cambridge, Mass.: MIT Press. 51–70.
- Todd, Peter M.; Gigerenzer, Gerd. 2000. 'Précis of Simple Heuristics That Make Us Smart'. *Behavioral and Brain Sciences* 23: 727–41.
- Wilson, Margo; Daley, Martin. 1992. 'The Man Who Mistook His Wife for a Chattel'. In: Barkow, Jerome H.; Cosmides, Leda; Tooby, John (eds). *The Adapted Mind*.

Evolution, Psychology, and the Generation of Culture. Oxford: Oxford University Press. 289–322.

Wollheim, Richard. 1999. *On the Emotions*. New Haven, Conn.: Yale University Press.

This page intentionally left blank