Jesse Prinz is associate professor of philosophy at the University of North Carolina, Chapel Hill. He is author of *Furnishing the Mind; Concepts and their Perceptual Basis and Emotional Perception*. He has held positions at the University of Maryland, Washington University in St. Louis, the California Institute of Technology, and the University of London.

Robert C. Solomon is Quincy Lee Centennial Professor of Business and Philosophy and Distinguished Teaching Professor at the University Texas at Austin. He is the author of *The Passions, In the Spirit of Hegel, About Love, A Passion for Justice, It's Good Business, Ethics and Excellence, Up the University, The Joy of Philosophy* and (with Kathleen M. Higgins) *A Short History of Philosophy* and *What Nietzsche Really Said*, and many other books.

Richard Wollheim taught at University College London from 1949 to 1981, and was Grote Professor of the Philosophy of Mind and Logic from 1963 onwards. He now teaches at the University of California Berkeley. He is the author of books in the philosophy of mind and the philosophy of art, including *Art and its Objects, Freud, The Thread of Life, Painting as an Art*, and *On the Emotions*. He is a Fellow of the British Academy and a Member of the American Academy of Arts and Sciences, and has been honoured by the International Psychoanalytical Association.

J. David Velleman is James B. and Grace J. Professor of Philosophy at the University of Michigan, Ann Arbor. His publications include *Practical Reflection* (Princeton University Press, 1989) and *The Possibility of Practical Reason* (Oxford University Press, 2000).

I. Emotions, Thoughts and Feelings: What is a 'Cognitive Theory' of the Emotions and Does it Neglect Affectivity?

ROBERT C. SOLOMON

I have been arguing, for almost thirty years now, that emotions have been unduly neglected in philosophy. Back in the seventies, it was an argument that attracted little sympathy. I have also been arguing that emotions are a ripe for philosophical analysis, a view that, as evidenced by the Manchester 2001 conference and a large number of excellent publications, has now become mainstream. My own analysis of emotion, first published in 1973, challenged the sharp divide between emotions and rationality, insisted that we reject the established notion that the emotions are involuntary, and argued, in a brief slogan, that 'emotions are judgments.' Since then, although the specific term 'judgment' has come under considerable fire and my voluntarist thesis continues to attract incredulousness the general approach I took to emotions has been widely accepted in both philosophy and the social sciences. When Paul Griffiths took on what he misleadingly characterized as 'propositional attitude' theories of emotion as the enemy of all that was true and scientifically worthy, I knew that we had made it.[1] Such ferocious abuse is surely a sign that we had shifted, in Kuhnian terms, from being revolutionary to becoming the 'normal' paradigm. The current counter-revolution of affect programmes and neuro-reductionism says a lot about who we are and how far we have come. (Progress in philosophy is moved more by this drama of one outrageous thesis after another—once called 'dialectic'—than by cautious, careful argument.)

The view that I represent is now generally referred to as the 'cognitive theory of emotions,' a borrowing from psychology and 'cognitive science.' The cognitive theory has become the touchstone of all philosophical theorizing about emotion, for or against. But what exactly is a 'cognitive' theory of emotions? The label 'cognitive theory' is not mine, and I fought it for years, not because

[1] P. Griffiths, *What Emotions Really Are* (Chicago, 1998).

it was wrong but because 'cognition' is so variously or ill-defined. In this talk, I would like to take on 'cognition' directly and try to say what I think it is and what it isn't, with particular reference to emotion. But to begin with, I want to reject, or at any rate call into question, the very *dimensions* of the emotional phenomena that are now under investigation. In recent work by Le Doux, Panksepp, and Damasio, for example, an emotion is sometimes presented as if it is more or less over and done with in 120 milliseconds, the rest being mere aftermath and cerebral embellishment. An emotion, so understood, is a preconscious, pre-cognitive, more or less automatic excitation of an affect programme. Now, I do not deny for a moment the fascinating work that these researchers have done and are doing, but I am interested, to put it polemically, in processes that last more than five minutes and have the potential to last five hours, five days, or five weeks, months, or even years. I am interested, in other words, not in those brief 'irruptive' disturbances but in the long-term narratives of Othello, Iago, Lily Bart and those of my less drama-ridden but nevertheless very emotional friends. I am interested in the meanings of life, not short-term neurological arousal.

Those bold and intriguing discoveries in the neurobiology of emotion have stimulated a mantra of sorts, 'emotion *before* cognition,' which rather leaves the cognitive theory, so to speak, with its pants down. (A fair turn around, one might argue, from my old slogan, 'emotions are judgments,' i.e., not Jamesian feelings or neurological events.) But the very statement of the new mantra provokes a cognitivist rejoinder: Surely the very fact of a *response* indicates some form of recognition, and (just to say the obvious) recognition is a form of cognition. What gets thrown into question, therefore, is not the intimate connection between emotion and cognition but the nature of cognition itself. Cognition is not to be understood only as conscious and articulate. There are primitive pre-conceptual forms of cognition, 'a cognitive neuroscience of emotion.'[2] These are not the forms of cognition or emotion that primarily interest me, perhaps, but they are extremely important in understanding not only the very brief phenomena studied by the neuroscientists but also the long-term emotional psycho-dramas that do interest me. Whatever else I may have meant or implied by my slogan 'emotions are judgments,' I was not thinking of necessarily conscious—and self-conscious—reflective, articulate judgments.

[2] R. Lane and L. Nadel, *The Cognitive Neuroscience of Emotion* (Oxford, 1999).

Emotions as 'Thoughts' and Other Things

'Cognition' is a not very informative technical term. It demands a translation into the vernacular. (If the charge against me is that I am stuck in what is now called, 'folk psychology,' I can live with that. Jerry Fodor may overstate the case when he insists that, 'folk psychology is the only game in town,' but it is certainly the Mother of All Games in Town.) The number of candidates that have been put forward to front the cognitive theory is impressive. Many authors, Jeffrey Murphy and Kendall Walton, for example, suggest *beliefs*. Jerome Neu, one of the prominent voices in the philosophy of emotions for more than twenty years, suggests that the cognitive elements that matter most are *thoughts*, a view that (at least nominally) goes back to Descartes and Spinoza.[3] Several philosophers (including myself) defend the theory that emotions are *evaluative judgments*, a view that can be traced back to the Stoics. Cheshire Calhoun has suggested 'seeing as' and Robert Roberts has offered us 'construal' as alternative, more perceptual ways of understanding cognition in emotion.[4] Other theorists, especially in psychology and cognitive science, play it safe with 'cognitive elements' or 'cognitive structures'.[5] Some psychologists split on the question of whether 'appraisals' are 'cognitions,' sometimes leading to a narrowed and critically vulnerable conception of both.[6] Many philosophers play it safe with the technical term 'intentionality,' although interpretations of this technical concept are often even less helpful than 'cognition.'[7] Pat Greenspan has played it coy with 'belief warrant' while rejecting the 'cognitive' theory in its more committal forms.[8] Michael Stocker is more directly combative when he rejects all of this in the defence of 'affect' and 'affective states,' although I have always suspected and will again here that Stocker's 'affect'

[3] Jerome Neu, *Emotion, Thought & Therapy* (Routledge, 1978).

[4] C. Calhoun, 'Cognitive Emotions?' in C. Calhoun and R. Solomon, *What is an Emotion?* (Oxford University Press, 1984); Robert Roberts, 'Propositions and Animal Emotion' *Philosophy* **71**, 147–56.

[5] E.g. A. Ortony, G. L. Clore and A. Collins, *The Cognitive Structure of Emotions* (Cambridge University Press, 1988); Robert Gordon, *The Structure of Emotions: Investigations in Cognitive Philosophy* (Cambridge University Press, 1987).

[6] R. Lazarus, J. Averill and E. Opton 'Towards a Cognitive Theory of Emotion', in *Feelings and Emotions*, Magda B. Arnold (Academic Press, 1970).

[7] A. Kenny, *Action, Emotion and Will* (London: Routledge, 1963).

[8] P. Greenspan, *Emotions and Reasons: An Inquiry into Emotional Justification* (New York: Routledge, 1988).

sneaks in a lot of what others portray as cognition.[9] Ronald De Sousa suggests 'paradigm scenarios,' an intriguing and more contextual and behavioural conception that is intended (among other things) to undermine the cognitive theory. (De Sousa, 1987)

Sometimes, the interpretation is absurdly more than the concept will bear, for example, in the overly committed conceptions of 'cognition' as *knowledge* (and therefore in some sense veridical). But it should be obvious that the cognition constituents of emotion can be wrong or mistaken. As my favourite philosophical author Nietzsche writes, 'The falseness of a judgment is not necessarily an objection to [it]. The question is to what extent it is life-promoting, life-preserving ...'[10] Whether or not the falseness of a cognition is an objection to an emotion (sometimes it is, sometimes it ain't), it is amply clear that whether or not it is an emotion or not is independent of its truth.

So, too, 'cognition' is interpreted in an overly narrow typically passionless cognitive science framework as 'information.' But while every emotion may presume information (for instance, in the recognition of its object) no amount of information (including information about one's own physiological and mental states) is sufficient to constitute an emotion. By the same reasoning I think the common linkage between emotion and belief is misleading. Beliefs and emotions are related in many important ways, belief as precondition or presupposition of emotion, and belief as brought about by emotion (say, by way of wishful thinking or rationalization).

Belief isn't the right sort of psychological entity to *constitute* emotion. Beliefs are necessarily dispositions, but an emotion is, at least in part, an *experience*. A belief as such is not ever experienced. Beliefs are propositional attitudes while many emotions are not (which is what's wrong with Griffiths's characterization). If Fred loves Mary and hates spinach, the objects of his emotions are Mary and spinach, respectively, not propositions. If Fred believes that spinach is good for you (and that, perhaps, is *why* he loves it) the object of his belief (but not his emotion) is the proposition that spinach is good for you.

Appraisal and evaluation or what Ortony *et al.* call 'valenced reactions' are necessary in emotion, even on the most basic neurological level, and belief too readily slides into the exclusively factual and epistemic if not into mere information. But an emotion is always

[9] M. Stocker, with E. Hegeman, *Valuing Emotions* (Cambridge University Press, 1996).

[10] F. Nietzsche, *Beyond Good and Evil* (New York: Random House, 1967).

value- or valence-laden.[11] Emotion as cognition does not point merely to information processing, and it cannot be captured in any list of beliefs or in terms of passionless states of knowledge.

Furthermore, there is considerable confusion concerning the 'level of awareness' of cognition, with neurological ('hard-wired') response at one end of the spectrum and then consciousness as recognition, as self-consciousness, as reflection, as articulation, and as deliberation at the other. The ambiguity of the word 'consciousness,' referring as it does both to unreflective awareness (the emotional experience) and to reflective self-consciousness (our recognition that we have such-and-such emotion), is the source of many problems, though I would argue that it is also the simple-minded dualism, based on the metaphor of 'reflection' (that is, mental activity versus the *observation* of that activity) that is at fault here. In the sense of consciousness as awareness, every emotion is (necessarily) conscious. In the sense of consciousness as articulate and self-conscious reflection, an emotion can become conscious only if one has (at the minimum) a language with which to 'label' it and articulate its constituent judgments. Thus I would challenge Jerome Neu's Blake-inspired title, 'A Tear is an Intellectual Thing,' on the grounds that it is not the *intellect* that is typically engaged in emotion. Thus I will also reject the view that cognitive theory—once distinguished from the intellect—*excludes affect*. The fact that many if not most emotions are non-reflective has no bearing on the question whether affect (so-called) might be an essential part of the cognitive aspect of emotional experience.

In his early work, and I see little evidence of radical change since, Jerome Neu took the defining element of emotion to be the very Spinozistic notion of a 'thought.' He makes it quite clear that one cannot have an emotion (or a particular kind of emotion) without certain types of thoughts. Emotions, simply stated, *are* thoughts, or dispositions to have thoughts, or defined by thoughts. (I am not considering here the very general Cartesian sense of '*cogitationes*' that would include virtually any mental process, state, or event, making the claim that emotions are thoughts utterly uninformative.) At the very least, Neu is correct when he says that thoughts are indicative of emotions and are produced during emotions.

I think that the notion of a 'thought' is too specific and involves too much intellect to provide a general account of the emotions. To be sure, a person with an emotion will have thoughts appropriate to the emotion and the context shaped and constrained by his or her

[11] A. Damasio, *Descartes' Error* (London: Macmillan, 1994).

language and culture. In the case of adult human emotions, I think that this may necessarily be so. But if belief is too dispositional to capture the essence of emotion, thoughts are too episodic for emotions, which often turn out to be enduring processes rather than mere episodes. Thus a thought may punctuate and manifest an emotion, but it is in itself not a process. *Thinking*, of course, is a process, but thinking is clearly too cerebral, too explicit, to characterize most emotions. A thought is a momentary appearance. It is a more or less articulate formation, and it is more or less independent of perception. Most thoughts involve words and the use of language, whether or not the thought is explicitly couched in words. Thus my thought of Paris (a postcard view of the Seine, looking towards Notre Dame) is a visual image but it's being a thought *of Paris* requires a complex act of recognition on my part. Thus I would say that dogs and babies may have emotions, perceptions and make judgments, but they do not have thoughts.

Philosophers since Frege confuse the matter by taking 'the thought' to be the proposition expressed by the thought, but the proposition alone (a logical construction) is never tantamount to a thought in the psychological sense, as an episodic phenomenon. Much less is a proposition (or a set of propositions) ever tantamount to an emotion. Thus the absurdity of Donald Davidson's much heralded analysis of emotion (following Hume's example of pride) in terms of a syllogism of propositions in logical sequence.[12] Philosophers also confuse the matter by conflating thoughts and thinking (Davidson, again), but although both might be involved in emotion (some emotions certainly 'get us thinking') it is *having* thoughts and having them without necessarily thinking that is most pronounced both as symptom and as constituent of emotion. When I have recurrent thoughts of violence or recurrent sexual fantasies a plausible hypothesis is that I have the appropriate (or rather, *in*appropriate) emotion. But insofar as thought is an aspect of emotion (rather than just a symptom or sign), it cannot merely be a proposition (or a set of propositions), and it must not be tied too tightly to the activity of thinking. (I would argue that it is also important not

[12] Donald Davidson (1977) 'Hume's Cognitive Theory of Pride' reprinted in Davidson (1980) *Essays on Actions an Events* (Oxford University Press), 277–90. Davidson's view was taken very seriously by many philosophers who never showed any interest in emotion, much less in any cognitive theory of emotion. But what gets left out of Davidson's reconstruction—as Hume himself clearly recognized—was pride, that is, the emotion. See Annette Baier, 'Hume's Analysis of Pride', *Journal of Philosophy*, **75** (1978), pp. 27–40.

to insist that thinking *cannot* be an aspect of emotion but rather only an antecedent or consequence of emotion.)

One feature of thoughts of particular interest to me which more or less follows from the distinction between thought and thinking is the fact that thoughts do not always appear by way of organized activity (like thinking) but rather appear in at least three ways, which I would summarize as 'conjured up' (when, for example, I think my way through a problem or try to remember the answer to a query), 'invited' (as when I work on a problem, give up on it for the evening, and the answer 'comes to me' in the middle of the night), and 'uninvited' (as when a thought 'pops' into my head, unwanted and unanticipated). This triple feature of thought is particularly relevant to the question whether and in what sense one can choose one's emotions for it is true both that one can (through thinking) choose one's thoughts and that thoughts can come unbidden. Insofar as thoughts are essential aspects of emotion one might note that thoughts are sometimes straightforwardly voluntary and even 'willed' (as in thinking), but thoughts also display considerable degrees of involuntariness, as when they 'pop' into my head (or, as Nietzsche wrote, 'A thought comes when *it* will, not when I will.')

Peter Goldie makes the interesting argument that while thoughts are voluntary, our imagination often 'runs away with us.' This depends on the nature of the distinction between thought and imagination. If a 'thought' is something abstract and merely conceptual (such as the *idea* that some one could possibly run off with my wife) while an image is by its very nature something fully fleshed and robust (such as an exquisitely detailed scenario in which my wife is having sex with another man) Goldie's claim is surely correct. But why should we restrict ourselves to such an emaciated sense of 'thought' or such an overly provocative sense of imagination? I think that Goldie is thinking primarily of thoughts 'conjured up' as opposed to thoughts merely invited or uninvited. I would say that both our thoughts and our imaginations are sometimes wilful, sometimes obsessive and beyond our control. Either way, wilful or obsessive, it is evidence that we have a strong emotion (whether or not we acknowledge it or know what it is) and it is suggestive of a sense in which our emotions are not in our control.

Beyond Belief

'Belief' has now become a catch-all term in cognitive science that specifies very little while it suggests something very specific. (Thus

emotion theorists in the late eighties, for instance Ronnie De Sousa and Robert Gordon, spent considerable time arguing that emotion cannot be captured by any combination of belief and desire but inevitably found that they were trying to get hold of a jellyfish.) Belief is too loosely tied to perception to account for those cases where one has an emotion immediately upon coming upon a situation, and it is too tightly tied to the logic of propositions to explain, for example, how it is that we can often hold conflicting (but not literally contradictory) emotions at the same time (what Patricia Greenspan raises as 'the problem of mixed emotions.')

Belief is typically described as a state, and though emotions may be states (that is, if they are of considerable duration and one ignores the dynamic engagement that goes on in emotion), it is surely inadequate to suggest that thus all emotions are states. That is why beliefs are often taken to be only 'cognitive preconditions' of emotion, not constitutive of emotion, since emotions are dynamic and often in flux while belief, as a holding onto a proposition, is a steady state. One either believes a proposition or not (although one might misleadingly express doubt or scepticism by saying that he or she 'sort of' believes that *p*.) Furthermore, beliefs are not experiences, though to be sure they shape and explain experiences. In Neu's vocabulary, they are always explanatory (they must always be postulated to explain behaviour and utterance in the third-person case) and not phenomenological. Belief may be perfectly appropriate in *explaining* emotion but it is inappropriate in the *analysis* of emotion.

These doubts about 'belief' explain the appeal of 'perception' as the 'cognitive element' most appropriate to the analysis of emotion. Ronnie De Sousa makes this case, as did John Dewey years ago, and I think that perception does indeed capture the heart of one kind of emotional experience, that which I would call 'immediate' (though without bringing in the heavy philosophical baggage that term conjures up in the history of epistemology). That is, in those examples where I have an emotional reaction to a situation unfolding right in front of my eyes, i.e. the sorts of examples employed (for obvious reasons) by William James in his classic analysis of emotion. Pointing out the close link between emotion and perception seems to me a plausible way of proceeding. Indeed one of its virtues is that it blocks the insidious distinction (still favoured by some positivistic psychologists) that perception is one thing, appraisal, evaluation, interpretation, and emotional response are all something else. Again, I prefer the concept of judgment precisely because it maintains these close ties to perception but at the same time, is fully conceivable apart from perception.

But when the trigger of an emotional response is a thought or a memory, the perception model looses its appeal. In general, when the object of emotion is something not immediately present, it makes little sense to say that the emotion is essentially a kind of perception. Take the appeal of such notions as 'construal' or 'seeing as.' Cheshire Calhoun defended 'seeing as' in criticizing my theory many years ago (in a book we co-edited).[13] As I have been revising my own 'judgment' theory over the years I have come more and more to construe 'judgment' as 'construal,' though I still think that 'judgment' has a number of advantages, not least of which is that it smacks less of reflection and is more pointedly less concerned with perception and other 'immediate' circumstances. 'Seeing as,' to be sure, is too tied to vision and thus perception, although (of course) it can be treated as a metonym (as Husserl, for instance, used the term) and extended to not only all of the senses but to all cognitive processing as well. But many of our emotions concern merely imaginary, distant, or abstract (but not therefore impersonal) concerns, and the 'seeing as' metonym is seriously stretched. Perhaps the point is better conceived in terms of 'construal,' a more consciously complex (as well as arguably voluntaristic) notion, but then I think the bias towards reflection cancels out these advantages.

Which brings me to Ronnie De Sousa's very fruitful idea of a 'paradigm scenario.' In his book, *The Rationality of Emotion*, De Sousa does not take this as a specification of cognition so much as an alternative to cognition. I have openly expressed my intrigue and admiration regarding this notion. Part of what is so exciting about it is that (unlike virtually all of the cognitive theories I have mentioned so far) it has an explicitly developmental and evolutionary bent. It takes a bold step in the direction of speculating how it is that we come to have the cognitions (or whatever) that constitute emotions, namely, by being taught to respond in certain ways (or taught what responses are appropriate) in specific situations. It thus has the virtue of being quite particularist, as opposed to those overly ambitious cognitive theories that try to draft broad generalizations that govern or constitute emotions. I would note that De Sousa as always been deeply involved in the theatre (and is pretty theatrical himself) and his theatrical shifting from emotion content to emotion context and behavioural training has always seemed to me a huge step forward in emotion research. It goes much further than superficially similar theories of 'action readiness' in that it postulates not only an ingredient in emotion and emotional experience but the

[13] C. Calhoun and R. Solomon, *What is an Emotion?* (Oxford University Press 1984).

dynamic of emotion as well. In what follows, I will also find two more virtues in De Sousa's theory, its explicit bringing in the body in a behavioural (not physiological) mode and its explicitly social nature, where other people are not just objects of our emotions or those who (in some sense) share our emotions but, in a critical sense, co-conspirators in the cultivation of our emotions.

Emotions as Judgments

Back there in ancient history, in 'Emotions and Choice' (1973) and *The Passions* (1976), I suggested 'judgment' to capture many of these insights. If Neu had the camaraderie of the neo-Stoical Spinoza, I could claim a linkage with the original Stoics, although I obviously rejected their conclusion that emotions as judgments are as such irrational. Briefly put, I take judgment in a way that is not episodic (although, to be sure, one can make a judgment at a particular moment). It is not necessarily articulate or for that matter conscious. (Neu clearly follows Freud in maintaining that thoughts, too, can be unconscious.) I take it as uncontroversial that animals make all sorts of judgments (e.g. whether something is worth eating, or worth chasing, or worth courting) but none of these are articulated or 'spelled out,' nor are they subject to reflection. We make non-reflective, non-deliberative, inarticulate judgments, for instance, kinaesthetic judgments, all the time. Kinaesthetic judgments are rarely deliberative and rarely merit conscious attention. Michael Stocker has a poignant story about his falling on the ice, thus making both his fear and his bodily awareness painfully conscious. But the example only illuminates the fact that such judgments are not usually conscious at all.

Judgments, unlike thoughts, are geared to perception and may apply directly to the situation we are in, but we can also make all sorts of judgments in the utter absence of any object of perception. Thus while I find the language of 'thought' just too intellectual, too sophisticated, and too demanding in terms of linguistic ability, articulation, and reflection, to apply to all emotions, judgment seems to me to have the range and flexibility to apply to everything from animal and infant emotions to the most sophisticated and complex adult human emotions such as jealousy, resentment, and moral indignation. In other words, I find the following to be essential features of emotion and judgment: they are episodic but possibly long-term as well. They must span the bridge between conscious and non-conscious awareness. They must accept as their

'objects' both propositions and perceptions. They must be appropriate both in the presence of their objects and in their absence. They must involve appraisals and evaluations without necessarily involving (or excluding) reflective appraisals and evaluations. They must stimulate thoughts and encourage beliefs (as well as being founded on beliefs) without themselves being nothing more than a thought or a belief. And (of considerable importance to me), they must artfully bridge the categories of the voluntary and the involuntary.

Thus emotions are like judgments. And emotions necessarily involve judgments. Does this entitle me to say that emotions *are* judgments? Well, not by logic alone, needless to say. But as a heuristic analysis and a way of understanding the peculiarities of emotion, I think so. But, of course, an emotion is not a single judgment. (In many traditional philosophical analyses, in Hobbes, Descartes, and Spinoza, the complex character of an emotion is reduced to a single one-liner.) An emotion is rather a complex of judgments and, sometimes, quite sophisticated judgments, such as judgments of responsibility (in shame, anger, and embarrassment) or judgments of comparative status (as in contempt and resentment). Emotions as judgments are not necessarily (or usually) conscious or deliberative or even articulate but we certainly *can* articulate, attend to, and deliberate regarding our emotions and emotion-judgments, and we do so whenever we think our way into an emotion, 'work ourselves up' to anger, or jealousy, or love. The judgments of love, for instance, are very much geared to the perceptions we have of our beloved, but they are also tied to all sorts of random thoughts, day-dreaming, hints and associations with the beloved, with all sorts of memories and intentions and imaginings. A judgment may be made at a certain time, in a certain place ('I loved you the first time I ever saw you') but one continues to make, sustain, reinforce, and augment such judgments over an open-ended amount of time.

I am willing to admit that different cognitive candidates may work better or worse for different emotions, and here I see further reason to heed and embellish the warning that Amelie Rorty and Paul Griffiths (for very different reasons) have issued, that 'emotion' is not an adequate category for cross-the-board analysis. Different emotions employ different kinds of cognition. This is the virtue, perhaps, of such non-committal notions as 'cognitive elements' or 'cognitive structures.' They are elastic enough to cover just about anything vaguely conceptual, evaluative, or perceptual. But while these seem to me to be useful conceptual tools for

working out the general framework of cognitive theory,[14] they clearly lack the phenomenological specificity that I am calling for here. Judgment seems to me to be, all in all, the most versatile candidate in the cognitive analysis of emotion. But by embracing (without distinction) the whole host of cognitive candidates, it is left open whether some emotions might be better analysed in terms of perception, others in terms of thoughts or judgments, others in terms of construals. The real work will have to be with regard to particular emotions, and often with specific regard for the particular instance of a particular sort of emotion.

What is Affect? Emotions, Feelings, and the Body

Michael Stocker and, more recently, Peter Goldie, have accused the cognitive theory of neglecting feelings, or 'affect.' I admit that in *The Passions* I was dismissive of the 'feeling theory' that then seemed to rule what passing interest there was in the emotions (particularly in the work of William James and his successors). I argued that whatever else it might be, an emotion was no mere feeling (interpreting this, as James did, as a bodily set of sensations). But what has increasingly concerned me ever since is the role of the body in emotion, and not only the brain. In my original theory, it was by no means clear that the body had *any* essential role in emotion. I presumed, of course, that all emotional experience had as its causal substratum various processes in the brain, but this had little to do with the nature of emotion as such, as experienced. But as for the various physiological disturbances and disruptions that serve such a central purpose in William James' analysis that the '*sensation IS the emotion*' (with all of the oomph that italics and caps can capture) and in later accounts of emotion as 'arousal,' I was as dismissive as could be, relegating all such phenomena to the causal margins of emotion, as merely accompaniments or secondary effects.

What has led me to this increasing concern about both the role of the body and the nature and role of feelings in emotion is in fact just the suspicion that my own cognitive theory had been cut too 'thin,' that in the pursuit of an alternative to the feeling theory I had veered too far in the other direction. I am now coming to appreciate that accounting for the feelings (not just sensations) in emotion is not a secondary concern and not independent of appreciating the essential role of the body in emotional experience. By this I do not

[14] Ortony *et. al.* (1988).

mean anything having to do with neurology or the tricky mind-body relationship linked with Descartes and Cartesianism but rather the concern about the kinds of *bodily experience* that typify emotion and the bodily manifestations of emotion in immediate expression. In retrospect, I am astounded that facial expression is hardly mentioned in *The Passions* (although, to be sure, my interest increased enormously when I met Paul Ekman some years later.) These are not mere incidentals. But understanding them will provide a concrete and phenomenologically rich account of emotional feelings in place of the fuzzy and ultimately content-less notion of affect.

The role of physiology in feeling is not straightforward. On the one hand, many physiological changes (including autonomic nervous system responses) have clearly experiential consequences, for instance flushing and the quickening of the heart beat. Many others (including most neurological activities) do not. James was rather indiscriminate in his specification of bodily and 'visceral' disturbances, but when he clearly referred to just those bodily processes (not necessarily disturbances) that had clear experiential or phenomenological effects, he did indeed capture something of what goes on in the *feeling* of emotion, although he short-changed the nature of the emotion itself. I now agree that feelings have been 'left out' of the cognitive account, but I also believe that 'cognition' or 'judgment' properly construed captures that missing ingredient. The analogy with kinaesthetic judgments suggests the possibility of bringing feelings of the body into the analysis of emotion in a straightforward way.

What are the feelings in emotion (though, to be sure, an emotion may last much longer than any given feelings, and feelings may outlast an emotion by several minutes or more)? The workings of the autonomic nervous system (quickened pulse, galvanic skin response, the release of hormones, sweating) have obvious phenomenological manifestations (feeling excited, 'tingly,' feeling flushed). Moreover, the whole range of bodily preparations and postures, many of them but not all of them within the realm of the voluntary, have phenomenological manifestations. Here too the well-catalogued realm of facial expression in emotion plays an important role. So do other forms of emotional expression. The category of 'action readiness' defended by Nico Frijda and others seems to me to be particularly significant here, not in terms of dispositional analysis of emotional behaviour but rather as an account of emotion feelings. Anger involves taking up a defensive posture. Some of the distinctive sensations of getting angry are the often subtle and

usually not noticed tensing of the various muscles of the body, particularly those involved in physical aggression. All of these are obviously akin to kinaesthetic feelings, the feelings through which we navigate and 'keep in touch with' our bodies. But these are not just feelings, not just sensations or perceptions of goings-on in the body. They are also *activities*, the activities of preparation and expression. The feelings of our 'making a face' in anger or disgust constitute an important element in our experiences of those emotions.

The voluntary status of these various emotion preparations and expressions is intriguing. Many gestures are obviously voluntary and the feelings that go along with them are the feelings of activity and not passivity. Many bodily preparations, even those that are not autonomic nervous system responses, are not voluntary and our feelings are more of the 'what's happening' sort than of 'I'm doing this.' Facial expressions are an especially intriguing category in this regard. Paul Ekman and others have analysed what most of us have recognized and that is the difference between (for example) smiles that are genuine (that is, to a certain extent involuntary) and smiles that are 'forced' (that is, voluntary but to some extent incompetent). Action readiness includes both autonomic (involuntary) as well as quite conscious and reflective posturing, for example, adopting a face and stance fit for the occasion, a darkened frown and threatening gesture in anger, a 'shame-faced' expression and a gesture of withdrawal or hiding in shame, a sentimental even teary-eyed smile and a tender gesture in love. And each of these has its phenomenological manifestations, its characteristic sensations or feelings that are part and parcel of emotional experience (whether noticed or recognized as such or not).

To put my current thinking in a nutshell, I think that a great deal of what is unhelpfully called 'affect' and 'affectivity' and is supposedly missing from cognitive accounts can be identified with the body, or what I will call (no doubt to howls of indignation) *the judgments of the body*. George Downing has put the matter quite beautifully in some of his recent work.[15] He writes of 'bodily micro-practices' and suggests that emotions are to a large extent constituted by these. This could, of course, be taken as just another attempt at behavioural reductionism, but Downing also insists that an emotion is essentially an experience. He also is quite happy to insist that cognitions (judgments) are also an essential part of any emotional experience. But he adds, and I have come to agree, that a good deal

[15] George Downing, 'Emotion Theory Revisited', in *Heidegger, Coping, and Cognitive Science: a Festschhrift for Hubert Dreyfus*, Vol. 2 (M.I.T. Press, 2001), pp. 245–70.

of cognition is of a radically pre-linguistic (very misleadingly called 'pre-cognitive') nature. Building on the work of Hubert Dreyfus and suggestions in Heidegger and Bourdieu, Downing insists that a good deal of emotional experience and even emotional knowledge can be identified in the development of these bodily micro-practices.

Does it make sense to call these judgments? I am sure the answer is yes, and I would defend this in two steps. First, I have already insisted that judgments are not necessarily articulate or conscious and so the sorts of discriminations we make and the construals that we perform are sometimes (often) made without our awareness of, much less reflection on, our doing so. Second, a relatively small store of human knowledge is of the form 'knowing that.' Philosophers, of course, are naturally concerned with such knowledge and that leads them not unnaturally to the prejudice that only such knowledge, propositional knowledge, is important. Not that they deny the need for all sorts of non-verbal skills of the 'knowing how' variety, but these are hardly the stuff of philosophical analysis, first, perhaps, because there may be nothing distinctively human about them (animals display such non-verbal skills at least as impressively as humans) and second, it is well-known that 'knowing how' cannot be reduced to any number of 'knowing that'-type propositions. But it is a distortion of cognition and consciousness to suggest that 'knowing that'-type propositional knowledge is in any way primary or independent of 'knowing how.' The thesis here obviously takes us back to Heidegger and Merleau-Ponty (and to a lesser extent, to Heidegger's one-time disciple Gilbert Ryle). But since I have already insisted that emotional judgments are not necessarily propositional the way is open to make the further claim that they are not necessarily 'knowing that' type cognitions either.

It goes without saying that many of our most 'knowing' responses to the world and the ways in which we bring meaning to our world have much more to do with the habits and practices we perform than the ways in which we think about and describe the world. Feelings of comfort (and discomfort) have a great deal to do with doing the familiar and finding ourselves acting in familiar ways with familiar responses. These feelings of comfort and discomfort range from felt-satisfaction, frustration and low-level anxiety to exuberant joy, full-blown anxiety, rage, and panic. Anger often involves feelings of discomfort, but to be anger (and not just frustration or irritation) the emotion must be further directed by way of some sort of blame, which in turn involves feelings of aggression and hostility, which may themselves be readily traced (as James did) to specific modes of arousal in the body (tensing of muscles, etc.). Shame is at

least in part a feeling of discomfort with other people, a feeling of rejection, as love is (in part) a feeling of unusual comfort with another. One might object that there is nothing *distinctively* bodily about any of this, and I would agree. But this is only to say that the Cartesian distinction of mind and body serves us ill in such cases, that it is only as embodied and mobile social beings that we have even the most primitive cognitions about the world to begin with. And more to the point it is in light of such pre-verbal and also active and engaged judgments that we have any emotions at all. We then embellish and enrich these through language, both by increasing (exponentially) the range of descriptions and behaviours and situations in which we become engaged (adding morality, aesthetics, and politics) and by increasing (logarithmically) the kinds of reactions we can have.

Thus the judgments that I claim are constitutive of emotion may be non-propositional and bodily as well as propositional and articulate, and they may further become reflective and self-conscious. What is cognition? I would still insist that it is basically judgment, both reflective and pre-reflective, both knowing how (as skills and practices) and knowing that (as propositional knowledge). A cognitive theory of emotion thus embodies what is often referred to as 'affect' and 'feeling' without dismissing these as unanalysable. But they are not analysable in the mode of conceptual analysis. That is what is right about Griffiths' otherwise wild charges. But neither are the feelings in question simply manifestations of the biological substratum, as James and Griffiths at least sometimes suggest. There are feelings, 'affects' if you like, critical to emotion. But they are not distinct from cognition or judgment and they are not mere 'read-outs' of processes going on in the body. They are judgments *of* the body, and this is the 'missing' element in the cognitive theory of emotions.

On Emotions and Choice

It was John Rawls who made me a radical. It was more than twenty-five years ago, when I was just starting to think my way through *The Passions*, that Rawls and I were having lunch while we were both visiting the University of Michigan. I explained my blooming thesis to him, and he asked, rather matter of factly, 'But surely when you say we choose our emotions you are saying something more than the fact that we choose what to do to bring about a certain emotion?' This was John Rawls, whose Great Book had just been published, and I was not about to say, 'Oh, well, yes, only that.'

Thus began a twenty-year stint of dramatic over-statement, to the effect of '*we choose our emotions*.'

There are two immediate obstacles to any such claim that emotions are matters of choice. The first is the obvious fact that emotions *seem* to happen to us, quite apart from our preferences or intentions. This phenomenological point is reinforced by a semantic-syntactic observation, that the language of the passions (starting with the word 'passion') is riddled with passivity, 'being struck by' and so on. (Though this set of observations should be balanced with another, that we sometimes feel guilty or glad about feeling what we feel, and that we often assess our emotions as warranted or not, wise or foolish, appropriate or inappropriate).

The second is the enormous range of emotions and emotional experiences, from being startled to carefully plotting one's revenge, from inexplicable panic upon seeing a little spider to a well-warranted fear of being audited by Internal Revenue, from falling 'desperately' in love to carefully cultivating a life-long loving relationship, from 'finding oneself' in a rage to righteous and well-considered indignation and a hatred of injustice. And it is not merely the difference between different emotions that is at stake here but (as several of the listed examples indicate) a difference in kinds of emotional experience in the same sort of emotion (fear, anger, love). The enormous range of emotions suggests that no single claim about choice will suit all emotions.

The voluntariness of emotions is obviously a contentious thesis that will require far more careful explication and defence than I can give it here. Let me limit myself to a few well-chosen arguments.

First of all, I certainly did not mean that emotions were *deliberate* actions, the results of overt plans or strategies. We do not think our way into most emotions. Nor do emotions fit the philosophical paradigm of intentional action, that is, actions that are preceded by intentions—combinations of explicit beliefs and desires and 'knowing what one is going to do.' Insofar as the emotions can be defended in terms of a kind of activity or action, it is not fully conscious intentional action that should be our paradigm. But between intentional and full-blown deliberate action and straightforward passivity—getting hit with a brick, suffering a heart attack or a seizure, for example, there is an enormous range of behaviours and 'undergoings' that might nevertheless be considered within the realm of the voluntary and as matters of responsibility.

Second, I was not claiming that having an emotion is or can be what Arthur Danto once called a 'basic action' (namely an action one performs *without performing any other action*, such as wiggling

one's little finger). One cannot 'simply' decide to have an emotion. One can, however, decide to do any number of things—enter into a situation, not take one's medication, think about a situation in a different way, 'set oneself up' for a fall—that will bring about the emotion. Or one might *act as if* one has an emotion, act angrily, for instance, from which genuine anger may follow. There is William James' always helpful advice: 'Smooth the brow, brighten the eye, contract the dorsal rather than the ventral aspect of the frame, and speak in a major key, pass the genial compliment, and your heart must be frigid indeed if it does not gradually thaw.' But this does not mean that we simply 'manipulate' or 'engineer' our emotions, as if *we* perform actions which affect or bring about *them*. Following Danto, one might say that virtually all human action—writing a letter, shooting a rifle, signalling a left-hand turn, working one's way through law school—involves doing something by doing something else, and this does not mean that the latter action *causes* the former. The one act (or course of action) *constitutes* the other.

Third, although it is certainly true that most of our emotions are not pre-meditated or deliberate, it is not as if *all* emotions are devoid of premeditation and deliberation. We often pursue love—the having of the emotion and not just the beloved—and we 'work ourselves into a rage,' at least sometimes with obvious objectives in mind (e.g. intimidating the other person). Whether an emotion is pre-meditated or deliberate, however, we may not experience the emotion as a choice among options. We may not think to ourselves, 'I could get angry now, or I could just resign myself to the fact that I'm a loser, or I could just forget it.' Given the situation, I simply choose to get angry. Nevertheless, I think that the notion 'choice,' like the notion of 'action,' is instructive here. It suggests a very different kind of framework for the study of emotion, one in which choice, intention, purpose, and responsibility play important if not central roles at least some if not most of the time. If we think of ourselves as authors of our emotions, we will reflect in such a way as to affect and possibly alter them. It would be nonsense to insist that, regarding our emotional lives, we are 'the captains of our fate,'[16] but nevertheless we are the oarsman and that is enough to hold that we are responsible for our emotions.

[16] The line is from William Henley's 'Invictus,' which has been forever tarnished by mass-murderer Timothy McVeigh, who quoted it immediately before his execution (June 2001).

II. The Emotions and their Philosophy of Mind

RICHARD WOLLHEIM

1. When I was invited by Yale University to deliver the Cassirer lectures,[1] I hesitated for a topic. I wanted something new. I proposed the emotions, and at that time my knowledge of the topic was so slight that I didn't know whether it was something that I had already written on or not.

I mention this fact because one thing that I have since learnt about the emotions is that such ignorance is in order. For it is one of those topics where grasping the extension of the term is inseparable from having some theory of the matter, however primitive. One way to explain this fact is to invoke the novelty of the term, for, in the sense in which it is used in this lecture, it is only about 300 years old. Another way, probably related, is to point to the fact that, not only are there belief *and* particular beliefs, desire *and* particular desires, but, when we refer to particular beliefs and to particular desires, we call them 'the belief that this' or 'the desire that that'. However there is no locution 'the emotion that this', or 'the emotion that that', which would indicate the presence of an emotion. It seems that ordinary language is an intermittent guide to the circumscription of the emotions.

However the uncertainty that I felt when I came to constructing a theory of the emotions went beyond anything I felt in demarcating the field of the emotions. For much of the time, I had the sense that I was engaged in work of pure improvisation. I tried to avoid first one rock, then another, and all the while I was aiming to keep as large a view as possible of the open sea: for without that, what is philosophy worth?

Now, somewhat more used to my own views, I see it less like that. There are, I now discern, certain basic ideas that, at one and the same time, organize my account of the emotions and correspond either to general features of the mind or to particular features of the emotions as I at any rate conceive of them. You can have a theory of the emotions—possibly even a true one—which disagrees with me

[1] Richard Wollheim, *On the Emotions: the Ernst Cassirer Lectures, 1991* (New Haven and London: Yale University Press, 1999).

IV. Emotion, Psychosemantics, and Embodied Appraisals

JESSE PRINZ

1. Two Theoretical Approaches to Emotion

There seem to be two kinds of emotion theorists in the world. Some work very hard to show that emotions are essentially cognitive states. Others resist this suggestion and insist that emotions are noncognitive. The debate has appeared in many forms in philosophy and psychology. It never seems to go away. The reason for this is simple. Emotions have properties that push in both directions, properties that make them seem quite smart and properties that make them seem quite dumb. They exemplify the base impulses of our animal nature while simultaneously branching out into the most human and humane reaches of our mental repertoires. Depending on where one looks, emotions can emerge as our simplest instincts or our subtlest achievements. This double nature makes emotions captivating, but also confounding. Researchers find themselves picking one side at the expense of the other, or packaging seemingly disparate components into unstable unions. I will defend a more integrative approach. For a more thorough treatment, see Prinz (forthcoming).

1.1 Noncognitive Theories

As I will use the terms, a cognitive theory of the emotions is a theory that maintains that all true emotions involve cognitions essentially. Noncognitive theories maintain that emotions do not necessarily involve cognitions. It is no easy matter to say what cognitions are. A failure to define this key term can easily lead to unproductive cross-talk. Despite that caveat, I will proceed without a definition. One can capture the difference between cognitive and noncognitive theories by considering some examples.

An especially simple form of noncognitive theory would be a pure feeling theory. Pure feeling theories identify emotions with qualitative feelings and nothing more. It is not clear whether any one has ever seriously defended such an account. In folk psychology, we sometimes employ a pure feeling theory of twinges and pangs. A

pure feeling theory of the emotions would regard emotions as analogous to these. Other relevant examples include the feeling of a buzz, glowing feelings, or unlocated pains. Freud is prone to describe emotions in this way. He insists that emotions cannot be unconscious because they are nothing but feelings (Freud, 1915). Hume (1739) can be read in this way as well, as when he insists that emotions do not represent things. A closer look at Hume, however, with his detailed taxonomy of emotion types, reveals a position that is far more sophisticated. Emotions are feelings, but they are individuated by the impressions and ideas that they have as causes and effects.

It is easier to find defenders of another class of noncognitive theories. In the 1880s William James and Carl Lange independently hit upon the suggestion that emotions are responses to patterned changes in the body. In this sense, emotions are embodied. For Lange (1885), emotions are principally responses to vascular changes. For James (1884), they are responses to more complex somatic states, including changes in skeletal muscles and visceral organs. Depth of inhalation, blood vessel dilation, heart rate acceleration, muscle tension, facial expression, and even instrumental actions can all factor into an emotional state. Fear might be an internal state that registers constricted vessels, blood flow to the extremities, a frowning open mouth, and flight behaviour. For James, the internal states are feelings, but they are not the *mere* feelings of a pure feeling theory. Emotions are feelings of the body. They are somatic feelings.

In recent times, Damasio (1994) has resuscitated the James-Lange theory, with a few alterations. Among the relevant bodily states Damasio now includes changes in the internal milieu, including changes of hormone levels. Damasio also denies that emotions are feelings, allowing that an unconscious state that registers a bodily change would qualify as an emotional response. And finally, Damasio argues that emotional responses can bypass the body. Our brains can respond as if our bodies had undergone a characteristic pattern of changes in the absence of such changes. For the brain, it can be 'as-if' the body had changed. This too would count as an emotion. James makes a similar claim in passing, but Damasio develops the idea much more extensively (1884: note 4).

Somatic theories enjoy considerable support. It is a commonplace that emotions are associated with actions, and the bodily response implicated by somatic theories can be viewed as response preparations. Increased blood flow in fear facilitates the flight or fright response. For James and Lange these changes are not consequences

of our emotions, but antecedents. Emotional feelings, at least, are feelings of the body preparing for action. To make this case, James and Lange both offer mental subtraction arguments. Imagine feeling an emotion as vividly as you can, and then subtract away each part of the feeling that owes to a bodily change. When the subtraction is complete, there is nothing left that would be recognized as the emotion.

Contemporary defences of the somatic approach emphasize empirical findings. Neural circuits that are associated with emotional response include structures that are independently associated with monitoring and maintaining bodily changes. Insular cortex and anterior cingulate cortex, for example, appear to be active in most functional neuroimaging studies of emotion (Damasio *et al.*, 2000). People with brain damage that prevents them from accurately monitoring bodily changes report a diminution of affective response (Critchley *et al.*, 2001).

These sources of evidence are suggestive, but far from decisive. A noncognitive theory must maintain that emotions are exhausted by noncognitive states. None of the evidence just mentioned rules out the possibility that the bodily concomitants of emotions may come along with cognitive states. Indeed, the evidence does not even show that emotions must have bodily components. Perhaps the evidence derives from sampling errors: placing too much emphasis on emotions that are especially primitive and intense. Even in these cases, a cognitive theorist could claim, for all the evidence thus far presented, that bodily responses are not components of emotions but mere accompaniments. The subtraction argument shows that bodily perturbations contribute to emotional feelings, but feelings may be contingent effects of emotions rather than essential features. Results from neuroimaging and self-reports from people with brain damage may, likewise, be picking up on emotional feelings. What pressure is there to think such feelings are constitutive of emotions? More to the point, what pressure is there to think that emotions can be exhaustively comprised by responses to bodily perturbations, be they felt or unfelt? Noncognitive theorists owe us more.

1.2 Cognitive Theories

The demand for further support is especially acute because noncognitive theories are seriously impoverished on the face of it. Emotions play central roles in our lives. They are ends (as when we seek pleasure, attachment, or amusement) and they are means (as when an emotion compels us to act). Emotions interact with

thought and reasons. Thinking about injustice can make a person angry and sadness can lead to thoughts about one's diminished prospects in life. Emotions also have intentional objects. One can be frustrated that *P*, afraid of *a*, delighted by *b*. In fact, emotions typically have intentional objects in two senses: particular and formal (Kenny, 1966). Each instance of an emotion is about some particular individual, situation, or event. Jones might be mad that her new camera is defective or mad that her husband is late again. In both cases, her anger has the same formal object; it concerns an offence against her. Sadness, in each normal instance, concerns loss, fear concerns danger, guilt concerns harmful transgressions. As Pitcher (1965) and others have pointed out, this contrasts markedly with twinges and pangs.

These kinds of considerations lead many to conclude that emotions are cognitive. One can easily explain why emotions interact with thoughts if one assumes that they *are* thoughts. Suppose, to take a simple view, that each emotion is comprised by a thought about some general property that bears on well-being. Anger may be the thought that there has been an offence against me. Sadness may be the thought that there has been a great loss. These thoughts directly explain why emotions have formal objects, because each explicitly refers to such an object. The particular objects of emotions are explained by combining thoughts. Suppose my dog Fido dies. I might first think that Fido is dead and then infer that this death is a loss. The inferred thought *constitutes*, on the simple cognitive view, the emotion.

Most cognitive views are not this simple. Solomon (1976) says that the judgment comprising an emotion cannot be separated from the judgment pertaining to the particular object. Anger that P is better rendered 'anger-that-p'. It is an evaluative judgment that construes an event as offence, rather than a reaction to an event that has been independently construed in a neutral way. Nussbaum (2001) says that having an emotion is a matter of assenting to a judgment that something important to personal well-being has transpired. Assenting can be regarded as a kind of judgment in its own right. In assenting, one evaluates a judgment pertaining to well-being as appropriate. If I feel sad, it is not just that I recognize a loss; I also judge that my sense of loss is warranted.

Both Solomon and Nussbaum contend that emotions can exist without any bodily concomitants. Their theories are purely cognitive. I come back to pure theories below. But first, I want to consider impure theories. Many cognitive theorists believe that emotions are thoughts plus some noncognitive component. One

might define emotions as evaluative judgments plus responses to bodily states. Appraisal theories in psychology are like this. Lazarus (1991) is a leading exponent (see also Arnold, 1960; Scherer 1984; Roseman, 1984). Emotions, he claims, involve feelings or action tendencies triggered by appraisal judgments. Each emotion involves the same appraisal 'dimensions.' There are six of these. We ask ourselves: has something relevant to my goals occurred? Is it congruent with goals? How is my ego involved? Who deserves credit or blame? What coping options are available? And What can I expect for the future? Emotions are distinguished by the different ways in which these questions can be answered. Anger involves the judgments that goals have been violated, that someone else is to blame, and that aggression is an available coping option. Every collection of answers can be summarized by what Lazarus calls a 'Core Relational Theme.' The appraisals constituting anger correspond to the theme that there has been a demeaning offence against me and mine. This is not an explicit judgment, but a way of capturing the gist of six more specific judgments answering to each dimension of appraisal.

Lazarus differs from Solomon and Nussbaum in allowing that emotions have noncognitive constituents. But, like them, he regards judgments essential. So we can ask all these researches the same question, Do all emotions necessarily involve judgments?

1.3 The Zajonc/Lazarus Debate

Zajonc (1980; 1984) is responsible for one of the most systematic critiques of cognitive theories in psychology. In his second article, Zajonc (1984) is especially concerned to refute Lazarus's theory, and he marshals several different kinds of arguments towards that end. Lazarus (1984) has responded to these arguments, and the resulting exchange has become a focal point in the battle between cognitive and noncognitive theories. I present some highlights.

In one line of argument, Zajonc contends that emotions are phylogenetically and ontogenetically prior to cognitions. Emotions are found in simpler animals, and they emerge before cognitions in human development. The difficulty with this contention is that we have no reason to deny that such creatures make judgments. Some of the concepts that figure into Lazarus's appraisal dimensions are quite sophisticated, including a concept of the self. But is it obvious that infraverbal creatures lack concepts such as danger or loss? If not, their analogues of fear and sadness may involve judgments that bear some kinship to our own. We should resist drawing *a priori*

conclusions about the kinds of judgments that infants and animals can form.

In another line of argument, Zajonc points out that a person can change her explicit appraisal judgment without changing her emotional state. Emotions can be recalcitrant. A person might continue to feel anger even after accepting the apology or believing the excuse of someone who offended her. Zajonc complains that Lazarus has not done enough to establish the link between emotion and appraisal.

In response, Lazarus (1984) points to decades of research showing that judgments can influence our emotional states. In an early series of studies, for example, Lazarus and colleagues (Speisman, *et al.*, 1965) induced different emotional responses to the same film-clip by altering the accompanying narrative. In more recent studies, he has obtained correlations between emotion labels and specific appraisal judgments (Smith and Lazarus, 1993, see discussion in Prinz, forthcoming). With regard to emotional recalcitrance, Lazarus has two available strategies. First, since he claims emotions contain noncognitive components, he can identify recalcitrant emotions with those whose accompanying feeling happen to out last the precipitating judgments. Second, he can claim that explicit changes in judgment do not always reverse incongruent unconscious judgments. There is ample independent evidence for this in social psychology. Once a false belief or prejudice has been planted, new evidence may fail to erase the initial judgment (Ross, *et al.*, 1975).

Zajonc calls on his own research in making a third argument against cognitive theories. He has been able to demonstrate a 'mere exposure effect' in preference formation. When subjects are briefly presented with unfamiliar stimuli (such as Chinese ideographs), they often perform at chance levels when given a subsequent recognition test. A previously presented stimulus may be judged as new. But preferences are effected by prior exposure. The more times a stimulus is presented, the more likely it is to be regarded favourably. For example, when American subjects were asked to speculate about which Chinese characters have a positive meaning, they were more likely to select the characters to which they had been most exposed, even if they had no explicit recall of seeing those characters (Zajonc, 1968). Zajonc draws two conclusions: subjects' judgments are informed by (possibly unconscious) affective responses, and those responses are noncognitive.

Both these conclusions are open to debate, but the later is especially relevant. If we grant that emotions factor into the mere exposure effect, should we conclude that emotions can occur

without cognition? We simply don't know. It is perfectly possible that subjects are performing unconscious appraisals. They may be unconsciously registering that the stimulus is novel. In our ancestral past, familiar stimuli that hadn't harmed us in the past would have been regarded as safer than entirely novel stimuli. Forced to choose between a familiar and a novel stimulus, subjects may form the appraisal that the familiar stimulus is more goal congruent, and a positive emotion will result. If this process is going on unconsciously, then the mere exposure effect is perfectly consistent with Lazarus's theory. Lazarus does not assume that his appraisals are conscious. To insist that appraisals must be conscious in order to count as cognitive would beg the question, construct a straw man, and depart from the orthodoxy within cognitive science (see, e.g., Nisbett and Wilson, 1977).

Zajonc's next batch of arguments is more convincing. He points to cases in which emotions are induced by direct physical means, as when the brain is stimulated by taking drugs, or reconfiguring facial muscles. The latter refer to the phenomenon of facial feedback. Making an emotional expression can give rise to the emotion itself even when one does not realize one is making an emotional expression (Zajonc, *et al.*, 1989). In one study, Strack, *et al.* (1988) asked subjects to fill out a questionnaire holding a pen in their mouths. Some subjects were asked to hold the pen between puckered lips and the other subjects were asked to hold it between their teeth with parted lips (conforming to a sour grimace and a smile-like facial configuration, respectively). In one part of the questionnaire subjects had to rate the amusement level of comic strips. Subjects in the teeth condition (who were unwittingly smiling) rated the comics as more amusing. Zajonc thinks there is a simple mechanism that mediates between facial change and emotional response. There is no need for the mediation of appraisals. Likewise, when we induce emotions through drugs.

Unfortunately, Lazarus (1984) does not offer a response to these kinds of cases. It would be desperate for him to propose that appraisals mediate facial feedback. There is no reason to think that appraisals are involved. There is nothing to appraise. We might take a smile as evidence that things are going well, but, in the Strack, *et al.* experiment, subjects do not even realize they are smiling. And, in any case, appraisals are unnecessary. Feedback effects could be achieved through direct wiring between brain states that register facial change and brain states that trigger the bodily responses associated with the corresponding emotion. There is one possibility available to Lazarus, however. Rather than postulating mediating

appraisals, he could deny that the feelings caused by facial feedback are emotions. He could say that they are recognisably *similar* to emotional feelings, but without appraisals, they fail to qualify as emotions themselves (see Clore, 1994). Likewise for feelings brought on by drugs. When probed about the giddy feeling brought on by a smile, one might respond, 'I feel as if I were happy, but I am not really happy; I have nothing to be happy about.'

To my mind, the most convincing evidence for emotions without cognitions comes from neuroanatomy. Zajonc argues that there are direct pathways from the most rudimentary perceptual centres to centres that initiate the bodily responses associated with an emotion. If so, those responses can begin before a person has had time to form an appraisal judgment. Zajonc (1984) mentions pathways from the retina to the hippocampus. These are no longer thought to be involved in emotions (they may be involved regulating the sleep cycle as a function of light), but other subcortical pathways exist. There is, for example a pathway from the superior colliculus and the pulvinar to the amygdala, which plays a central role in mediating between perception and the physiological aspects of some emotions. The superior colliculus and pulvinar are very rudimentary perceptual structures that convey information to the amygdala before the neocortex has gotten involved. Appraisal judgments of the kind Lazarus imagines—judgments involving such concepts as 'ego' and 'loss'—are likely to be implemented in the neocortex. If emotional somatic responses can be induced before the neocortex comes on line, then Lazarus is wrong to claim that appraisals must precede those responses.

On the face of it, the anatomical evidence could be dismissed in the same way that I suggested Lazarus respond to facial feedback. Perhaps responses caused by the subcortical pathway to the amygdala should not qualify as true emotions. Perhaps they are just emotion-like. This objection is considerably less plausible here. In the case of facial feedback, one is tempted to say that the responses are not true emotions, because they are not playing typical emotion roles. In contrast, subcortically induced emotions are often quite typical. LeDoux (1996) has argued that the subcortical path to the amygdala underlies speedy fear responses to simple stimuli. Imagine seeing a snake. Before you can recognize it as such, your earliest perceptual centres have discerned the characteristic coiled shape and they have sent emotion centres into action. It is a shortcut that allows us to save time in situations where time is of the essence. Now why call this an emotional response? The reason has to do with the stimulus. As in facial feedback, the end-state feels like

an emotion. But here, the elicitor is a paradigmatic emotion elicitor. Snakes are dangerous. And snakes are not the only stimulus that can travel the subcortical path. Sudden noises, angry faces, sudden loss of support, creeping bugs, looming objects, and total darkness, are among the many things that may be able to spark an emotion without the cortical assistance. If you find yourself in a state that feels just like fear after seeing something that really is threatening, there is no reason to deny that you are really experiencing fear. All this happens without the need for appraisals. Appraisals might come into the picture after the bodily response. But this won't help Lazarus. His theory requires that appraisals come first. If a state that plays the right emotion role can be initiated without appraisal, then appraisals are not essential for some emotions.

Lazarus tries to dismiss Zajonc's appeal to neuroanatomy on the grounds that our interpretation of the brain depends on our psychological theories. Perhaps there are subcortical appraisals. It is good to exercise caution when arguing about the brain, but the caution is misplaced here. There is no reason to locate appraisals in the superior colliculus or the amygdala. These structures are fairly well understood. We know how their cells respond and how they are wired. No viable interpretation could assign the concepts comprising Lazarus's dimensional appraisals or core themes to the networks in these parts. The colliculus responds to raw perceptual signals and the amygdala serves as a bridge between these and structures that control basic bodily states. The brain shows how we can move from an image to a racing heart without bringing in concepts of ego, blame, expectancy, or goals. The simplest perception can trigger the bodily perturbations that we experience as emotions.

1.4 The Emotion Problem

The preceding two subsections ended with contradictory results. First we saw that noncognitive theories are explanatorily anaemic. They cannot explain the rich interactions between emotions and reasoning, nor can they account for the two senses in which emotions can be said to have intentional objects. In a word, noncognitive theories fail to capture the fact that emotions are meaningful.

Cognitive theories are well suited to capture the meaningfulness of emotions. They identify emotions with judgments or with more complex mental episodes that include judgments as parts. But there is empirical evidence that emotions can arise in the absence of judgments. Elementary perceptions of external stimuli can send us

reeling without cognitive mediation. And when this occurs, the emotion seems no less meaningful than it would in other cases. Fear caused by seeing a snake lunge towards you is surely significant. It is not like an undirected and inexplicable pang.

So we have a serious puzzle. The fact that emotions are meaningful, reason sensitive, and intentional suggests that they must be cognitive. The fact that some emotions arise without the intervention of the neocortex suggests that emotions cannot *all* be cognitive. The emotions that arise in this way seem to be meaningful. This suggests that being meaningful does not require being cognitive. Noncognitive states are explanatorily anaemic and cognitive states are explanatorily superfluous. Noncognitive theories give us too little, and cognitive theories give us too much. Call this the Emotion Problem.

2. Embodied Appraisals

2.1 Psychosemantics

The Emotion Problem is essentially a problem about getting meaning on the cheap. To solve it, we need a way of showing how emotions can have the semantic properties that they seem to have without claiming that emotions are judgments. If we are seek out a explanation of how mental states can have semantic properties without being judgments, we do not need to look very far. Prevailing theories of intentionality that have been developed within the philosophy of mind are well suited to this end. These theories were not devised to explain the emotions. They were devised to explain how concepts refer. If such theories do a reasonable job with concepts, then they may apply to mental states quite broadly. If they help explain the semantic properties of the emotions, then we may have an independently motivated solution to the Emotion Problem.

I think that informational theories are especially promising (Dretske, 1981; Fodor, 1990). These theories begin with the principle that representation involves law-like dependencies. A mental state refers to things that would cause that state to be tokened. Dretske (1988) combines this idea with a teleological component. Mental states refer to those reliable causes that they have the function of detecting. More succinctly, a mental state refers to what it is set up to be set off by. The second condition (being set off) captures the informational component of the theory. If a mental state is reliably set off by some class of things, then it carries the information that some item in the class is present (in much the way

the smoke carries the information that a fire is present). But this condition is too permissive on its own. Many things cause our mental states to be tokened. It would be a mistake to say that a mental state refers to anything that causes it. This would make error impossible. The first condition (being set up) is recruited to narrow down the content. All mental states are acquired in some way, usually by inheritance or learning. Of the many things that reliably cause a mental state to be tokened only those things for which it was initially generated fall within its extension. Dretske likes to make his case by appeal to simple artefacts. If you light a match near a smoke detector it will beep, but it was set up to detect smoke caused by fires, not the flames from a match. Likewise, a concept of water refers to water, even though it is occasionally set off by encounters with other clear liquids. It was created in the context of water detection.

It is beyond the scope of this discussion to defend informational semantics. I will assume that some version will work for concepts (see Prinz, 2002). Here, I want to show that it also explains how emotions get their contents. According to Jamesian theories, emotions are the internal states that register bodily changes. On the face of it, these states represent bodily changes if they represent anything at all. This is not inconsistent with informational semantics. Such states are reliably set off by patterned changes in the body. But is it their function to detect such changes? Why did we develop minds that detect *patterned* bodily changes? Why do body-pattern detecting states get set up and why do they persist? An obvious answer is that these patterns happen to occur under conditions that are important to us. The patterns associated with fear (such as flight preparation or freezing) happen to occur when we are facing immediate physical dangers. Danger is, thus, another reliable cause of the inner states that registers fleeing or freezing patterns. And it is a cause that has especially good claim to being the one for which such states are attained in the first place. We come to be good body-pattern detectors (through evolution and learning), because body patterns co-occur with matters of grave concern. States that *register* body changes may *represent* the more abstract relational properties that induce those changes in us.

2.2 A Theory of Emotion

This suggests the following theory of emotions. Emotions are, as James suggested, inner responses to bodily changes. They are, in that sense, embodied. But emotions represent matters of concern.

They represent things like danger and loss—the core relational themes emphasized by cognitive theorists. The embodied states represent core themes because they have the function of being reliably caused by core themes. If we define an appraisal as any mental state that represents an organism-environment relation that bears on well-being, then the embodied states in question qualify as appraisals. They are embodied appraisals.

Notice that embodied appraisals do not *describe* the states that they represent. One can have an embodied state that represents an immediate physical danger without having concepts of immediacy, physicality, or danger. This is just a consequence of informational semantics. Such theories do not require highly structured representations. The beep emitted by smoke detector might be said to represent 'smoke from fire here now,' but it does not decompose into meaningful sub-beeps. It is semantically primitive. Complex contents do not need complex representations. Defenders of cognitive theories assume that emotions can only designate core relational themes if emotions are judgments, thoughts, or some other kind of concept-laden, structured states. This simply isn't true. To represent appraisal core relational themes, emotions need only occur, reliably, when those themes occur.

Cognitive theorists might raise an objection at this point. Surely emotions can reliably co-occur with core relational themes only if they contain judgments. How else could they reliably coincide with dangers, losses, offences, and all the rest? LeDoux's snake case points towards an answer. When we see a snake, our bodies enter into characteristic patterns that are registered by the embodied states that I have identified with emotions. There is no judgment involved, just a snake image in the early visual system. Now suppose that the same body pattern is innately triggered by several other kinds of images as well, such as bugs, looming objects, darkness, threatening faces, and blood. Suppose these things cause the same body pattern via direct links from perception to body control centres. The state that registers that body pattern is reliably caused by each of these things, but they have this common effect in virtue of the fact that they all instantiated a common property. They are all dangers. The embodied state is a danger detector, because danger is the property that gets the items in this hodgepodge to have an impact. Snakes, spiders, and darkness cause our hearts to race and palms to sweat in virtue of the fact that they were hazards to our ancestors.

In this initial state, judgments play no role. Danger detection is entirely noncognitive. Later we may come to recognize that bodily

patterns in question are occurring in dangerous situations. We may acquire a danger concept, and that concept may come to be deployed in situations that would be hard to recognize by casual observation (e.g., judging that the newly elected politician is dangerous). The old body response can come to have a broader range of application in this way. After concepts are acquired, fear can be triggered by cognitive appraisals. But these cases are derivative. Explicit judgments come after a meaningful emotion already exists. Because such judgments are inessential to emotions, we should not count them as emotion constituents even when they do occur. They are no more a part of an emotion than a premise is a part of the conclusion it supports. They are causes, not components.

2.3 Solving the Emotion Problem

The embodied appraisal theory offers a solution to Emotion Problem. With noncognitive theories, it says that emotions are embodied, and it denies that judgments are needed for emotion elicitation. With cognitive theories, it says that emotions represent core relational themes. This helps explain why emotions interact with thinking. If emotions represent core themes, thoughts pertaining to those themes will be rationally tied to emotions. Thoughts that provide evidence that one is in danger warrant fear, and fear warrants thoughts about strategies for coping with danger.

Emotions can be said to have intentional objects in the two required senses on the embodied appraisal theory. It is a central theme of the theory that emotions have formal objects. These are just the core relational themes that emotions have the function of reliably detecting. It would take more work to show how emotions attain particular objects. How does my fear that it will rain get connected up with my thought that it may rain? Answering this question fully would require more detail than I can provide here, but I can present the basic strategy. The emotion and the thought can be linked in three ways. Semantically, the thought that it will rain may be taken as a potential danger, which may serve to induce state of fear as an entailment. Syntactically, the fear state may be bound to the thought in whatever way that mental representations are generally bound. Imagine hearing a voice coming from a moving face. We somehow link these together. Perhaps the same method of linking (or some other independently motivated method) can bind emotions to thoughts about particular objects. Finally, there may be a counterfactual dependence between the emotion and the thought such that the emotion would not have occurred if the thought had

not. Syntactic links and counterfactual dependencies could link emotions to thoughts even if emotions had no meaning. The semantic point adds to the story by explaining why emotions get linked to representations of particular objects. This simply wouldn't make sense if emotions were meaningless sensations.

In sum, the embodied appraisal theory explains how emotions bridge the gap between thoughts and feelings. They are structurally simple embodied states, but they carry the kind of information that full-blown cognitions can carry. Cognitive theories have been right about content, and noncognitive theories have been right about form.

3. Generalizing the Account

By way of conclusion, I want to consider whether the embodied appraisal theory can account for the full range of emotions we experience. Cognitive theorists sometimes claim that some emotions have no bodily concomitants. Call these disembodied emotions. Cognitive theorists also claim that some emotions must have cognitive components, in addition to any bodily components they might comprise. I briefly consider these objections in turn (see Prinz, forthcoming, for more). Faced with such objections, one might simply concede that emotions do not form a coherent class (cf. Griffiths, 1997). My hope is to show that emotions are unified. All cases can be explained in terms of embodied appraisals.

Disembodied emotions include calm passions, such as loneliness or aesthetic appreciation, and long-standing emotions, such as the enduring love one feels for a spouse. Equating emotions with brief bodily perturbations does not capture these cases. Therefore, the embodied appraisal theory does not generalize.

Alleged disembodied emotions can be handled in one of three ways. Some are not emotions at all. Harré (1996) uses loneliness as an example, but it is not obvious that loneliness is an emotion. It certainly isn't a paradigm case. A theory of emotions should begin with clear cases, and then provide a way of determining whether less clear cases qualify as emotions. The embodied appraisal theory says that loneliness is an emotion only if it represents a core relational theme *and* is embodied. If it isn't embodied, we can rule that it isn't an emotion. This would only beg the question if loneliness were a clear case. The same can be said of calm aesthetic responses and what might be termed 'polite passions' (as when one says, 'I am sorry I couldn't make it to the reception'). Other

putative disembodied emotions turn out to be embodied. They involve bodily changes that are harder to detect than cases involving sympathetic responses of the autonomic nervous. If loneliness is an emotion, it probably involves a reduction of heart rate rather than an increase. Loneliness is presumably related to mild sadness, which bears such bodily marks.

The third strategy for handling disembodied emotions is to draw a distinction between dispositional and occurrent states. Most mental state types have both dispositional and occurrent forms. One can say, 'wool makes Jones itchy,' and 'wool is making Jones itchy right now.' Or, 'Jones believes that Quine was right about analyticity,' and 'Jones is using that belief right now in her reasoning.' Likewise, I regard long-standing emotions as dispositions. That I love my spouse all the time is an enduring disposition to have occurrent states of love (compare 'I detest country music' or 'I am disgusted by floral wall paper' or 'I am outraged by cinema violence'). An occurrent state of love is an embodied reaction of the kind one has when one encounters the object of one's love. In line with this, Bartels and Zeki (2000) showed brain activation in limbic areas associated with bodily response when subjects viewed pictures of their lovers. I would add that long-standing love does not count as love *unless* it carries a disposition to such embodied states. If someone says, 'I love my spouse, but I never experience flutters or giddiness or cuddly tenderness in relation to him' we would doubt her sincerity. As with itchiness, standing emotions are parasitic on their embodied manifestations.

The final objection to the embodied appraisal theory concedes that all emotions contain embodied appraisals (at least dispositionally), but asserts that some emotions contain cognitive elements as well. Certain emotions pertain to situations that can only be appreciated by creatures with command of very advanced concepts. Emotions such as jealousy, guilt, shame, and indignation come to mind. These implicate ideas of infidelity, transgression, the self, and justice. They may be uniquely human. The natural explanation of such emotions aligns with cognitive theories. Guilt, for example, might be said to involve the evaluative judgment that I have committed a transgression that wrongfully harms someone whose well-being matters to me. Unlike fear, which might be triggered by a simple percept, guilt seems to require conceptual mediation.

I can only gesture towards an account of these cases here. Theorists from Descartes to Ekman have proposed that some emotions are basic. Other emotions blend these basic emotions

together (contempt may be anger plus disgust; thrills may be joy plus fear) and others elaborate basic emotions by bringing in other resources. Oatley and Johnson-Laird (1997) talk of elaborated emotions, which are basic emotions plus thoughts. Schadenfreude might be joy plus the thought that someone is miserable. Guilt may be sadness plus thoughts about my transgressions. Romantic jealousy may be a blend of sadness, fear, anger, and disgust plus thoughts of a lover's infidelity.

I think this approach is almost right, but I would add an important modification. Rather than viewing the cognitive elaborations as constituent parts of an emotion, I regard them as calibrating causes. A calibrating cause is a mental state that triggers an emotion under a specific set of conditions that are somewhat different than the conditions that elicited the emotion initially. Calibrating causes are external to emotions, and may be highly variable. Rather than having one single thought that triggers guilt or jealousy, we may have rich calibration files, containing many thoughts and perceptions. Jealousy can be sparked by the judgment that my lover has been unfaithful, or by the thought that she has been arriving home late, or by the smell of another's cologne on her shirt, or by the glance she makes towards a passing stranger. Guilt can be caused by the explicit judgment that I have transgressed or by recognizing that I have done something specific that, on a prior occasion, I recognized as a transgression. These disparate triggers all serve to align my embodied state with core relational themes. My jealousy calibration file places a blended embodied state under the nomic control of signs of infidelity, and that blend thereby comes to represent infidelity when it is triggered by *any* item in the file. Jealousy does not contain its calibrating causes. Which one would it contain? Jealousy represent infidelity because it is set up to be set off by infidelity, and the items in a calibration file contribute to that. But these items are no more a part of jealousy than are the eyes by which the jealous person perceives infidelity's palpable signs.

If these suggestions are right, then the embodied appraisal theory is truly general. All emotions are nothing more than embodied appraisals or dispositions to embodied appraisals. All are structurally and semantically analogous. All have embodied form and appraisal content. Embodied appraisal theory offers unity in two senses. It binds all emotions into a coherent category, and it reconciles the differences that have pushed cognitive and noncognitive theorists into such opposite directions.

Bibliography

Arnold, M. B. 1960. *Emotion and Personality* (New York, NY: Columbia University Press).

Bartels, A. and Zeki, S. 2000. 'The neural basis of romantic love', *NeuroReport*, **11**, 3829–34.

Clore, G. L. 1994. 'Why emotions require cognition', In P. Ekman and R. Davidson (eds) *The nature of emotion: Fundamental questions* (Oxford: Oxford University Press).

Critchley, H. D., Mathias, C. J. and Dolan, R. J. 2001. 'Neural correlates of first and second-order representation of bodily states', *Nature Neuroscience* 2001; **4**, 207–12.

Damasio, A. R. 1994. *Descartes' Error: Emotion Reason and the Human Brain* (New York, NY: Gossett/Putnam).

Damasio, A. R., Grabowski, T. J., Bechara, A., Damasio, H., Ponto, L. L. B., Parvizi, J. and Hichwa, R. D. 2000 'Subcortical and Cortical Brain Activity During the Feeling of Self-generated Emotions', *Nature Neuroscience*, **3**, 1049–56.

Dretske, F. 1981. *Knowledge and the Flow of Information* (Cambridge, MA, MIT Press).

Dretske, F. 1988. *Explaining Behavior* (Cambridge, MA, MIT Press).

Fodor, J. A. 1990. 'A Theory of Content, I & II', In *A Theory of Content and Other Essays* (Cambridge, MA: MIT Press).

Freud, S. 1915. 'The unconscious', In *The Standard Edition of the Complete Works of Sigmund Freud*, volume 14, James Strachey (trans.) (London: Hogarth Press).

Griffiths, P. 1997. *What emotions really are* (Chicago: University of Chicago Press).

Harré, R. 1986. 'The Social Constructivist Viewpoint', In R. Harré (ed.) *The Social Construction of Emotions* (2–14). Oxford: Blackwell.

Hume, D. 1739/1978. *A treatise of human nature*. Nidditch, P. H. (ed.) (Oxford: Oxford University Press).

James, W. 1884. 'What is an Emotion?', *Mind*, **9**, 188–205.

James, W. 1894. 'The Physical Basis of Emotion', *Psychological Review*, **1**, 516–29.

Kenny, A. 1963. *Action, Emotion and Will* (London: Routledge and Kegan Paul).

Lange, C. G. 1885. *Om sindsbevaegelser: et psyko-fysiologisk studie*. Kjbenhavn: Jacob Lunds. Reprinted in *The Emotions*, C. G. Lange and W. James (eds), I. A. Haupt (trans.) (Baltimore: Williams & Wilkins Company 1922).

Lazarus, R. 1984. On the primacy of cognition. *American Psychologist*, **39**, 124–29.

Lazarus, R. S. 1991. *Emotion and Adaptation*. (New York: Oxford University Press).

LeDoux J. E. 1996. *The Emotional Brain* (New York, NY: Simon & Schuster).

Nisbett, R. E. and Wilson, T. D. 1977. 'Telling more than we can know: Verbal reports on mental processes', *Psychological Review*, **84**, 231–59.
Nussbaum, M. 2001. *Upheavals of thought* (Oxford: Oxford University Press).
Oatley, K. and P. N. Johnson-Laird 1987. 'Towards a cognitive theory of emotions', *Emotions and Cognition*, **1**, 29–50.
Pitcher, G. 1965. 'Emotion', *Mind*, **74**, 324–46.
Prinz, J. J. 2002. *Furnishing the Mind: Concepts and Their Perceptual Basis* (Cambridge, MA: MIT Press).
Prinz, J. J. (forthcoming). *Emotional Perception* (New York: Oxford University Press).
Roseman I. J. 1984. 'Cognitive Determinants of Emotion: A Structural Theory', In P. Shaver (ed.) *Review of Personality and Social Psychology, Volume 5* (11–36). (Beverly Hills, CA: Sage).
Ross, L., Lepper, M. and Hubbard, M. 1975. 'Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm', *Journal of Personality and Social Psychology*, 32
Scherer, K. R. 1984. 'On the Nature and Function of Emotion: A Component Process Approach', In K. R. Scherer and P. Ekman (eds) *Approaches to Emotion* (293–318). (Hillsdale, NJ: Erlbaum).
Smith, C. A. and Lazarus, R. S. 1993. 'Appraisal components, core relational themes, and the emotions', *Cognition and Emotion*, **7**, 233–69.
Solomon, R. 1976. *The Passions: Emotions and the meaning of life, Indianapolis* (IN: Hackett Publishing Company).
Speisman, J. C., Lazarus, R. S., Mordkoff, A. M. and Davison, L. A. 1964. 'The experimental reduction of stress based on ego-defense theory', *Journal of Abnormal and Social Psychology*, **68**, 367–80.
Strack, F., Martin, L.L. and Stepper, S. 1988. 'Inhibiting and facilitating conditions of facial expressions. A nonobtrusive test of the facial feedback hypothesis', *Journal of Personality and Social Psychology*, **54**, 768–77.
Zajonc, R. B. 1968. 'Attitudinal effects of mere exposure', *Journal of Personality and Social Psychology Monograph Supplement*, **9**, 1–27.
Zajonc, R. B. 1980. Feeling and thinking: Preferences need no inferences. *American Psychologist*, **35**, 151–75.
Zajonc, R. B. 1984. 'On the Primacy of Affect', *American Psychologist*, **39**, 117–23.
Zajonc, R. B., Murphy, S. T. and Inglehart, M. 1989. 'Feeling and Facial Reference: Implications of the Vascular Theory of Emotion', *Psychological Review*, **96**, 395–416.

V. Emotions and the Problem of Other Minds

HANNA PICKARD

Can consideration of the emotions help to solve the problem of other minds? Intuitively, it should. We often think of emotions as public: as observable in the body, face, and voice of others. Perhaps you can simply see another's disgust or anger, say, in her demeanour and expression; or hear the sadness clearly in his voice. Publicity of mind, meanwhile, is just what is demanded by some solutions to the problem. But what does this demand amount to, and do emotions actually meet it? This paper has three parts. First, I consider the nature of the problem of other minds. Second, I consider the publicity of emotions. And third, I bring these together to show how emotions can help to solve the problem.

Traditionally, there are two problems of other minds: one epistemological, one conceptual. The epistemological problem asks how you can know, or how you can be justified in believing, that another person has a mind at all: that there exist other subjects of experience. The conceptual problem asks how you can so much as understand that there could exist other minds or subjects of experience: how you can have the concept of another's mind or experience. But why suppose these problems exist? They both arise, in part, from the same idea. This is the idea of an ontological distinction between experience (mind) and behaviour (body): they are not the same type of thing.

The idea is intuitive. Consider, for instance, the possibility of pretence. Another person may be carrying on as if in pain, say, grimacing and crying out, but not be in pain at all. Yet you might be wholly taken in by her performance. There may be no discernible difference between real and pretend behaviour. Now pretence is not, nor perhaps could it ever be, the norm. But what it brings to the fore is the possibility of a discrepancy between experience and behaviour. If you can know how another is behaving, yet mistake her experience, the most obvious explanation is that her behaviour is observable, while her experience is hidden from view. Thus the distinction between the two.

The result of this distinction is that the only experience you can

reasoning can use certain principles typical of emotional reasoning, but emotional reasoning can seldom use intellectual principles while still remaining in the emotional mode. Thus, taking a broad, detached intellectual perspective typically eliminates the emotional experience; however, taking the narrow and involved perspective typical of emotions does not necessarily eliminate the intellectual mode.

As both types of logic are useful in different circumstances, it is to the benefit of each of us to integrate them in an optimal manner. Although such integration is difficult, it is possible to achieve it, with various levels of success.

References

Ben-Ze'ev, A. 1993. *The Perceptual System: A Philosophical and Psychological Perspective* (New York: Peter Lang).

Ben-Ze'ev, A. 2000. *The Subtlety of Smotions* (Cambridge, Mass.: MIT Press).

Ben-Ze'ev, A. (forthcoming). 'Emotion as a subtle mental mode'. In R. Solomon (ed.) *Thinking about Feeling: Contemporary Philosophers on Emotion* (Oxford: Oxford University Press, forthcoming).

Bergson, H. 1907. *Creative Evolution* (New York: Holt, 1911).

Frijda, N. H. 1988. 'The laws of emotion', *American Psychologist*, **43**, 349–58.

Margolis, H. 1987. *Patterns, Thinking, and Cognition: A Theory of Judgment* (Chicago: The University of Chicago Press).

Nussbaum, M. C. 2001. *Upheavals of Thought: The Intelligence of Emotions* (Cambridge: Cambridge University Press).

Redding, P. 1999. *The Logic of Affect* (Ithaca: Cornell University Press).

Rock, I. 1983. *The Logic of Perception* (Cambridge, Mass.: MIT Press).

Smith, C. A. and Kirby, L. D. 2000. 'Consequences require antecedents: Toward a process model of emotion elicitation', In J. Forgas (ed.), *Feeling and Thinking: The Role of Affect in Social Cognition* (pp. 83–106) (Cambridge: Cambridge University Press).

Spinoza, B. (1677). *Ethics*. In E. Curley (ed.), *The Collected Works of Spinoza* (Princeton: Princeton University Press, 1985).

X. Emotion and Desire in Self-Deception

ALFRED R. MELE

According to a traditional view of self-deception, the phenomenon is an intrapersonal analogue of stereotypical interpersonal deception.[1] In the latter case, deceivers *intentionally* deceive others into believing something, *p*, and there is a time at which the deceivers believe that *p* is false while their victims falsely believe that *p* is true. If self-deception is properly understood on this model, self-deceivers intentionally deceive themselves into believing something, *p*, and there is a time at which they believe that *p* is false while also believing that *p* is true.

Elsewhere (most recently in Mele, 2001), I have criticized the traditional conception of self-deception and defended an alternative, deflationary view according to which self-deception does not entail any of the following: intentionally deceiving oneself; intending (or trying) to deceive oneself, or to make it easier for oneself to believe something; concurrently believing each of two contradictory propositions. Indeed, I have argued that garden-variety instances of self-deception do not include any of these things. On my view, to put it simply, people enter self-deception in acquiring a belief that *p* if and only if *p* is false and they acquire the belief in *a suitably biased way*.[2] Obviously, this shoulders me with the burden of showing

[1] This tradition is embraced in influential work on self-deception in philosophy, psychology, psychiatry, and biology. See, e.g., Pears, 1984, Quattrone and Tversky, 1984, Gur and Sackeim, 1979, and Trivers, 1985. *Stereotypical* interpersonal deception does not exhaust interpersonal deception.

[2] Two points should be made. First, I have never defended a statement of *necessary* and sufficient conditions of entering self-deception in acquiring a belief that *p*, but only statements of characteristic and jointly sufficient conditions. (For a recent statement, see Mele, 2001, pp. 50–1.) Second, the requirement that *p* be false is purely semantic. By definition, one is *deceived in* believing that *p* only if *p* is *false*; the same is true of being *self-deceived in* believing that *p*. The requirement does not imply that *p*'s being false has special importance for the *dynamics* of self-deception. Biased treatment of data may sometimes result in someone's believing an improbable proposition, *p*, that happens to be *true*. There may be self-deception in such a case, but the person is not self-deceived in believing

what suitable bias amounts to, and I have had a lot to say about that. The suitability at issue is a matter of kind of bias, degree of bias, and the nondeviance of causal connections between biasing processes (or events) and the acquisition of the belief that *p*.[3]

In Mele, 2001 (pp. 106–12), I suggested a test for relevant bias. I called it 'the impartial observer test,' and I argued that its appropriateness is underwritten by the ordinary concept of self-deception. Here is an improved version: If S is self-deceived in believing that p, and D is the collection of relevant data readily available to S, then if D were made readily available to S's impartial cognitive peers (including merely hypothetical people) and they were to engage in at least as much reflection on the issue as S does and at least a moderate amount of reflection, those who conclude that p is false would significantly outnumber those who conclude that p is true.[4] This is a test for the satisfaction of a necessary condition of being self-deceived in believing that p. One requirement for impartiality in the present context is that one neither desire that p nor desire that ~p. Another is that one not prefer avoidance of either of the following errors over the other: falsely believing that p and falsely believing that ~p. The kind of bias at issue may broadly be termed 'motivational or emotional bias.' Although I have discussed biasing causes and processes—especially motivational ones—at length, I have left it open that a motivationally biased treatment of data is not required for self-deception and that emotions sometimes do the biasing work without motivation's playing a biasing role. This is one of the two possibilities that I explore in this essay. The other is a more moderate thesis about the place of emotion in self-deception.

1. Background: Biased Belief and Self-Deception

In the present section, after briefly describing some mechanisms relevant to the production of motivationally biased belief of a sort

[3] On deviant and nondeviant causation in this connection, see Mele, 2001, pp. 121–23.

[4] Cf. Mele, 2001, p. 106. The improvement is the 'reflection' clause. An issue may be so boring to one's impartial cognitive peers that they do not reflect on it and reach no conclusion about it.

that *p*, nor in acquiring the belief that *p*. On a relevant difference between being deceived *in* believing that *p* and being deceived *into* believing that *p*, see Mele, 1987, pp. 127–8.

appropriate to self-deception, I sketch a general account of such belief. My primary aim is to prepare the way for my discussion of emotionally biased belief in Section 2. I have reviewed empirical evidence of motivationally biased belief elsewhere (most recently in Mele, 2001) and I will not do so again here.

Attention to some phenomena that have been argued to be sources of unmotivated or 'cold' biased belief sheds light on motivationally biased belief. A number of such sources have been identified in the psychological literature, including the following two.

1. Vividness of information. A datum's vividness for a person often is a function of the person's interests, the concreteness of the datum, its 'imagery-provoking' power, or its sensory, temporal, or spatial proximity (Nisbett and Ross, 1980, p. 45). Vivid data are more likely to be recognized, attended to, and recalled than pallid data. Consequently, vivid data tend to have a disproportional influence on the formation and retention of beliefs.

2. The confirmation bias. People testing a hypothesis tend to search (in memory and the world) more often for confirming than for disconfirming instances and to recognize the former more readily (Baron, 1988, pp. 259–65; Klayman and Ha, 1987; Nisbett and Ross, pp. 181–82). This is true even when the hypothesis is only a tentative one (as opposed, e.g., to a belief one has). People also tend to interpret relatively neutral data as supporting a hypothesis they are testing (Trope *et al.*, 1997, p. 115).

Although sources of biased belief apparently can function independently of motivation, they may also be triggered and sustained by desires in the production of particular motivationally biased beliefs.[5] For example, desires can enhance the vividness or salience of data. Data that count in favour of the truth of a proposition that one hopes is true may be rendered more vivid or salient by one's recognition that they so count. Similarly, desires can influence which hypotheses occur to one and affect the salience of available hypotheses, thereby setting the stage for the confirmation bias.[6] Owing to a desire that p, one may test the hypothesis that p is true

[5] I develop this idea in Mele, 1987, ch. 10 and Mele, 2001. Kunda, 1990 develops the same theme, concentrating on evidence that motivation sometimes primes the confirmation bias. Also see Kunda, 1999, ch. 6.

[6] For motivational interpretations of the confirmation bias, see Friedrich, 1993 and Trope and Liberman, 1996, pp. 252–65.

rather than the contrary hypothesis. In these ways and others, a desire that p may contribute to the acquisition of an unwarranted belief that p.

Sometimes we generate our own hypotheses, and sometimes others suggest hypotheses to us—including extremely unpleasant ones. If we were consistently to concentrate primarily on confirmatory instances of hypotheses we are testing, independently of what is at stake, that would indicate the presence of a cognitive tendency or disposition that uniformly operates independently of desires. For example, it would indicate that desires never play a role in influencing the proportion of attention we give to evidence for the falsity of a hypothesis. However, there is powerful evidence that the 'confirmation bias' is much less rigid than this. For example, in one study (Gigerenzer and Hug, 1992), two groups of subjects are asked to test 'social-contract rules such as "If someone stays overnight in the cabin, then that person must bring along a bundle of firewood ..."' (Friedrich, 1993, p. 313). The group asked to adopt 'the perspective of a cabin guard monitoring compliance' showed an 'extremely high frequency' of testing for disconfirming instances. The other group, asked to 'take the perspective of a visitor trying to determine' whether firewood was supplied by visitors or a local club, displayed the common confirmation bias.

An interesting recent theory of lay hypothesis testing is designed, in part, to account for motivationally biased belief. I examined it in Mele, 2001, where I offered grounds for caution and moderation and argued that a qualified version is plausible.[7] I named it the 'FTL theory,' after the authors of the essays on which I primarily drew, Friedrich, 1993 and Trope and Liberman, 1996. I will offer a thumbnail sketch of the theory shortly. First, an explicit application of it to self-deception should be noted.

On James Friedrich's PEDMIN—'primary error detection and minimization'—model of lay hypothesis testing, 'detection and minimization of crucial errors is ... the central organizing principle' in this sphere (1993, p. 299). Regarding self-deception, Friedrich writes:

> a prime candidate for primary error of concern is believing as true something that leads [one] to mistakenly criticize [oneself] or lower [one's] self-esteem. Such costs are generally highly salient and are paid for immediately in terms of psychological discomfort. When there are few costs associated with errors of self-deception (incorrectly preserving or enhancing one's self-image), mistakenly revising one's self-image downward or failing to boost it appropriately should be the focal error. (p. 314)

[7] See Mele, 2001, pp. 31–49, 63–70, 90–91, 96–98, 112–18.

The basic idea of the FTL theory is that lay hypothesis testing is driven by a concern to minimize making costly errors. The errors in question are false beliefs. The cost of a false belief is the cost, including missed opportunities for gains, that it would be reasonable for the person to expect the belief—if false—to have, given his desires and beliefs, if he were to have expectations about such things. A central element of the FTL theory is the notion of a 'confidence threshold'—or a 'threshold,' for short. The lower the threshold, the thinner the evidence sufficient for reaching it. Two thresholds are relevant to each hypothesis: 'The acceptance threshold is the minimum confidence in the truth of a hypothesis,' p, sufficient for producing a belief that p; and 'the rejection threshold is the minimum confidence in the untruth of a hypothesis,' p, sufficient for producing a belief that ~p (Trope and Liberman, 1996, p. 253). Acquiring the belief terminates hypothesis testing. The two thresholds often are not equally high, and the acceptance and rejection thresholds respectively depend 'primarily' on 'the cost of false acceptance relative to the cost of information' and 'the cost of false rejection relative to the cost of information.' The 'cost of information' is simply the 'resources and effort' required for gathering and processing 'hypothesis-relevant information' (p. 252).

Confidence thresholds are determined by the strength of desires to avoid specific costly errors together with information costs. Setting aside the latter costs, the stronger one's desire to avoid falsely believing that p, the higher one's threshold for belief that p. These desires influence belief in two ways. First, because, other things being equal, lower thresholds are easier to reach than higher ones, belief that ~p is a more likely outcome than belief that p, other things being equal, in a hypothesis tester who has a higher acceptance threshold for p than for ~p. Second, the desires at issue influence how we test hypotheses, not just when we stop testing them (owing to our having reached a relevant threshold). Recall the study in which subjects asked to adopt 'the perspective of a cabin guard' showed an 'extremely high frequency' of testing for disconfirming instances whereas subjects asked to 'take the perspective of a visitor' showed the common confirmation bias.

It might be claimed that if avoidance desires of the kind under discussion function in the second way, they function in conjunction with beliefs to the effect that testing-behaviour of a specific kind

will tend to help one avoid making the costly errors at issue. It might be claimed, accordingly, that the pertinent testing-behaviour is performed for a reason constituted by the desire and an instrumental belief of the kind just mentioned and that this behaviour is therefore performed with the intention of trying to avoid, the pertinent error. The thrust of these claims is that the FTL theory accommodates the confirmation bias, for example, by invoking a model of intentional action.

This is not a feature of the FTL model, as its proponents understand it. Friedrich claims that desires to avoid specific errors can trigger and sustain 'automatic test strategies' (p. 313), which supposedly happens in roughly the nonintentional way in which a desire that p enhances the vividness of evidence for p. A person's having a stronger desire to avoid falsely believing that ~p than to avoid falsely believing that p may have the effect that he primarily seeks evidence for p, is more attentive to such evidence than to evidence for ~p, and interprets relatively neutral data as supporting p, without this effect's being mediated by a belief that such behaviour is conducive to avoiding the former error. The stronger desire may simply frame the topic in a way that triggers and sustains these manifestations of the confirmation bias without the assistance of a belief that behaviour of this kind is a means of avoiding a certain error. Similarly, having a stronger desire that runs in the opposite direction may result in a sceptical approach to hypothesis testing that in no way depends on a belief to the effect that an approach of this kind will increase the probability of avoiding the costlier error. Given the stronger desire, sceptical testing is predictable independently of the agent's believing that a particular testing style will decrease the probability of making a certain error. So at least I have argued elsewhere (Mele, 2001, pp. 41–49, 61–67).

I will not defend this thesis again here. Nor am I claiming that the FTL theory is acceptable without qualification. The theory may accurately describe what happens in some or many cases of lay hypothesis testing that results in belief, and in many or all cases of self-deception.

One more piece of background is in order. Elsewhere, I have distinguished between 'straight' and 'twisted' self-deception (Mele, 1997b; 1999; 2000; 2001, pp. 4–5, 94–118). In straight instances, we are self-deceived in believing something that we want to be true. In twisted instances, we are self-deceived in believing something that we want to be false (and do not also want to be true). Twisted self-deception may be exemplified by an insecure, jealous husband who believes that his wife is having an affair despite possessing only

relatively weak evidence for that proposition and unambivalently wanting it to be false that she is so engaged.[8]

The FTL theory applies straightforwardly to twisted self-deception. Whereas, for many people, it may be more important to avoid acquiring the false belief that their spouses are having affairs than to avoid acquiring the false belief that they are not so engaged, the converse may well be true of some insecure, jealous people. The belief that one's spouse is unfaithful tends to cause significant psychological discomfort. Even so, avoiding falsely believing that their spouses are faithful may be so important to some people that they test the relevant hypothesis in ways that, other things being equal, are less likely to lead to a false belief in their spouses' fidelity than to a false belief in their spouses' infidelity. Furthermore, data suggestive of infidelity may be especially salient for these people. Don Sharpsteen and Lee Kirkpatrick observe that 'the jealousy complex'—that is, 'the thoughts, feelings, and behaviour typically associated with jealousy episodes'—can be regarded as a mechanism 'for maintaining close relationships' and appears to be 'triggered by separation, or the threat of separation, from attachment figures' (1997, p. 627). It certainly is conceivable that, given a certain psychological profile, a strong desire to maintain one's relationship with one's spouse plays a role in rendering the potential error of falsely believing one's spouse to be innocent of infidelity a 'costly' error, in the FTL sense, and more costly than the error of falsely believing one's spouse to be guilty. After all, the former error may reduce the probability that one takes steps to protect the relationship against an intruder. The FTL theory provides a basis for a plausible account of twisted self-deception (see Mele, 1999 and 2001, ch. 5).

2. Emotions in Self-Deception

I turn to possible roles for emotions in self-deception. Insofar as emotions are causes of belief-biasing desires or are partially constituted by such desires, they have a clear bearing on self-deception, if the FTL theory is on the right track. Consider Bob, who is self-deceived in believing that his wife, Ann, is not having an affair. Bob's love for Ann, or his fear that he cannot get along without her, may be a partial cause of his desire that she is not having an affair

[8] On this case, see Barnes, 1997, ch. 3; Dalgleish, 1997, p. 110; Lazar, 1999, pp. 274–77; and Pears, 1984, pp. 42–4. Also see Mele, 1987, pp. 114–18.

and, thereby, of his being self-deceived about this. If that desire increases the salience of his apparent evidence of her fidelity or helps shape his relevant confidence thresholds, emotions that contribute to the desire play an indirect part in this. Furthermore, Bob may fear that Ann is guilty of infidelity, and if a constituent of his fear is a desire that she is innocent, then the role the desire plays in his self-deception may be attributed to the fear, insofar as the fear encompasses the desire.[9]

There are additional possibilities. Obviously, people are averse to anxious feelings. Such feelings may be caused by reflection. In some cases, a desire that one's anxiety subside may play a role in attenuating or halting reflection on an unpleasant hypothesis, thereby decreasing the likelihood of one's undermining a contrary hypothesis to which one is attracted. Anxiety is an emotion: here, again, an emotion has an indirect role in potential self-deception. Some emotions may also help to explain some instances of self-deception by weakening one's motivation to assess evidence carefully (see Forgas, 1995, p. 50), thereby increasing the probability that one's beliefs will be unduly influenced by one's desires. Grief may do this.

Do emotions figure more prominently in some cases of self-deception? In the remainder of this essay, I explore two hypotheses about this.[10]

1. The solo emotion hypothesis. In some instances of entering self-deception in acquiring a belief, an emotion, but no desire, makes a biasing contribution to the production of that belief.

2. The direct emotion hypothesis. In some instances of entering self-deception in acquiring a belief, an emotion makes a biasing contribution to the production of that belief that is neither made by a desire nor causally mediated by a desire.[11]

My primary aim is to convey a sense of what may be said for and against these hypotheses and of difficulties involved in investigating

[9] Fear that $\sim p$ is plausibly understood as being partly constituted by desire that p. See, e.g., Davis, 1988.

[10] Elsewhere (Mele, 2000, pp. 125–29), I criticized a third hypothesis, what I called 'the anxiety reduction hypothesis'—the thesis that the function of self-deception is to reduce present anxiety (Barnes, 1997, Johnston, 1988; cf. Tesser *et al.*, 1989).

[11] To simplify discussion, I formulated both hypotheses in terms of entering self-deception in acquiring a belief. Entering self-deception in retaining a belief and remaining in self-deception in continuing to believe something also require attention.

them. As I will explain, the second, more modest hypothesis is plausible, and our knowledge about emotion is too thin to warrant a confident rejection or endorsement of the first hypothesis.

Regarding an instance of twisted self-deception, Tim Dalgleish writes: 'it is inappropriate to suggest that jealous persons desire or are motivated to find that their partners are unfaithful; rather, their emotional state is priming the relevant processing systems to gather evidence in a biased fashion' (1997, p. 110). Dalgleish's contention is that, in cases of this kind, emotion plays biasing roles of the sort I attributed to desires in straight self-deception. For example, jealousy may prime the confirmation bias by prompting a jealous man to test the hypothesis that his wife is unfaithful, and it may increase the salience of apparent evidence of infidelity. There is evidence that emotions operate in these ways. As Douglas Derryberry reports, there is evidence that 'emotional states facilitate the processing of congruent stimuli' and that 'attentional processes are involved in [this] effect' (1988, pp. 36, 38), and Gordon Bower and Joseph Forgas review evidence that emotions make 'emotionally congruent interpretations of ambiguous stimuli more available' (2000, p. 106).[12] For example, Jed's jealousy may make him highly attentive to rare memories of Jane's seemingly being flirtatious or secretive and help generate jealousy-congruent interpretations of relatively neutral data.

The jealous Jed scenario is unlikely to confirm the solo emotion hypothesis. Sharpsteen and Kirkpatrick, suggest, plausibly, that 'the jealousy complex' is 'a manifestation of motives reflecting both sexual and attachment concerns' (1997, p. 638). Jealousy is intimately bound up with desires that jealous people have concerning their relationships with the people of whom they are jealous. It is a truism that indifference about one's relationship with a person precludes being jealous of that person.[13] (Being envious of someone is another matter.) Indeed, it is plausible that if a desire for close romantic attachment is not a constituent of paradigmatic romantic jealousy, it is at least a significant, partial cause of such jealousy. If this plausible proposition is true, then if Jed's being jealous of Jane affects the hypotheses he frames about her, the vividness of his

[12] Reviews of the 'mood congruence effect' include Bower and Forgas, 2000 and Forgas, 1995.

[13] If a woman is jealous because her date is flirting with another woman, is she jealous of her date or the other woman? Ronald de Sousa expresses the proper usage succinctly: 'the person one is jealous *of* plays an entirely different part in one's jealousy from that of the rival *because of whom* one is jealous' (1987, p. 75).

evidence, and the focus of his attention, it is very likely that an attachment desire plays a biasing role.[14]

Again, 'the jealousy complex' can be regarded as a mechanism 'for maintaining close relationships' and appears to be 'triggered by separation, or the threat of separation, from attachment figures' (Sharpsteen and Kirkpatrick, p. 627). This suggests that the effects of jealousy are partly explained by a desire for the maintenance of a close relationship. That desire may be at work in Jed's biased cognition. The desire may contribute to Jed's having a stronger desire to avoid falsely believing that Jane is faithful than to avoid falsely believing that she is unfaithful and, accordingly, contribute to his having a lower acceptance threshold for the hypothesis that she is having an affair than for the contrary hypothesis. The desire, given its psychological context, including, importantly, the jealousy associated with it, may also help enhance the salience of evidence of threats to the maintenance of his relationship with Jane, help prime the confirmation bias in a way favouring the belief that she is having an affair, and so on.

Defending the solo emotion hypothesis is a challenging project. Owing to the tight connection between emotions and associated desires, testing empirically for cases of self-deception in which emotion, and not desire, biases belief is difficult. Constructing compelling conceptual tests also is challenging. For example, if all emotions, or all emotions that might plausibly bias beliefs, are partly constituted by desires, it is difficult to show that there are beliefs that are biased by an emotion, or by some feature of an emotion, but not at all by desires, including desires that are constituents of the biasing emotions. Even if there is a conceptual connection between types of emotions and types of desires as partial causes, rather than between types of emotions and types of desires as constituents, it would have to be shown that emotions sometimes contribute to instances of self-deception to which the desires involved in producing the emotions make no belief-biasing contribution. Furthermore, even if some emotions are neither partially constituted nor partially caused by a relevant desire (typical instances of surprise are like this), the solo emotion hypothesis requires that such an emotion's biasing contribution to self-deception not be causally mediated by a desire either and, more generally, that the emotion not contribute to

[14] The conjunction of '*x* affects *y*' and '*z* is a constituent or a cause of *x*' does not entail '*z* affects *y*.' The brake pedal on Smith's car is a constituent of his car and his car affected Jones. But the brake pedal did not. Smith's car fell on Jones as he was repairing a flat tire. That explains the qualification 'very likely' in the text.

self-deception in combination with any biasing desire. I will return to this hypothesis shortly.

The direct emotion hypothesis is more modest.[15] Perhaps in some or many instances of self-deception, biasing roles are played both by (aspects of) emotions and by desires that are intimately related to the biasing emotions—as part to whole, or as a partial cause or effect, or as responses to the emotions (as in the case of a desire to be rid of one's present anxiety).[16] In some such cases, some biasing roles played by emotions may be direct, in the relevant sense. Perhaps an emotion can prime the confirmation bias or enhance the salience of emotion-congruent data without doing so simply in virtue of a constituent desire's playing this role and without the effect's being causally mediated by a desire. One who knows only of Jed's evidence for and against the proposition that Jane is having an affair and of his desire for the maintenance of a close relationship with her is hard put to understand why Jed believes that this proposition is true. People with much stronger evidence of infidelity than Jed has often believe that their spouses are innocent of infidelity, even though they, like Jed, strongly desire the maintenance of close relationships with their spouses. Indeed, some common philosophical examples of straight self-deception feature such people. The information that Jed is jealous helps us understand why he believes what he does. His jealousy is an important, instructive part of the psychological context in which he acquires his infidelity belief. Perhaps Jed's jealousy plays a role in the production of his biased belief that is not played by the pertinent desire alone.

Consider another scenario. Ed is angry at Don for a recent offence. His anger may prime the confirmation bias by suggesting an emotion-congruent hypothesis about Don's current behaviour—for example, that Don is behaving spitefully again. Ed's anger may also increase the salience of data that seem to support that hypothesis. There is evidence that anger tends to focus attention selectively on explanations in terms of 'agency,' as opposed to situational factors (Kilter *et al.*, 1993). Perhaps Ed's anger leads him to view certain aspects of Don's behaviour as more goal-directed and more indicative of a hostile intention than he otherwise would. If anger has a desire as a constituent, it is, roughly, a desire to lash out against the target of one's anger. Possibly, anger can play the biasing roles

[15] More specifically, although, necessarily, any emotion that makes a solo biasing contribution to self-deception makes a direct biasing contribution (in the pertinent sense of 'direct'), an emotion that makes a direct biasing contribution might not make a solo one.

[16] The categories of effect and response are not mutually exclusive.

just mentioned without any constituent desire's playing them and in the absence of causal mediation by a desire.

If an emotion can play a direct biasing role in self-deception, perhaps an emotion may contribute to an instance of self-deception that involves no desires as significant biasing causes. Perhaps the solo emotion hypothesis is true, despite the challenges it faces. It is conceivable, perhaps, that Ed enters self-deception in acquiring the belief that Don is behaving spitefully now, that the process that results in this belief features his anger's playing the biasing roles just described, and that no desires of Ed's have a biasing effect in this case. Now, on the assumption that Ed believes that Don is behaving spitefully despite having stronger evidence for the falsity of that hypothesis than for its truth, an FTL theorist will find it natural to suppose that Ed had a lower threshold for acceptance of that hypothesis than for rejection of it, that the difference in thresholds is explained at least partly in terms of relevant desires, and that this difference helps to explain Ed's acquiring the belief he does. But this supposition is open to debate, and I will not try to settle the issue here.

I mentioned that testing the solo emotion hypothesis empirically would be difficult. This point helps to explain the limited scope that Joseph Forgas claims for his 'affect infusion model' of the effects of affective states on social judgments (1995; cf. Bower and Forgas 2000).[17] Sketching some background will enable me to say how. Forgas identifies two 'mechanisms of affect-infusion: affect-priming and affect-as-information' (p. 40). The former is a matter of the 'selective influence [of affective states] on attention, encoding, retrieval, and associative processes' during substantive information processing (p. 40). In a nice illustration of the latter, 'when subjects were asked to make off-the-cuff evaluative judgments about their happiness and life satisfaction through a telephone survey, their responses were significantly different depending on whether they were feeling good (interviewed on a pleasant, sunny day) or feeling bad (interviewed on a rainy, overcast day). Once their attention was called to the source of their mood (the weather), however, the mood effects were constrained' (p. 53).[18] Commenting on such studies, Norbert Schwarz writes: 'rather than computing a judgment on the basis of ... features of a target, individuals may ... ask themselves "How do I feel about it" [and] in doing so, they may mistake [certain] feelings ... as a reaction to the target' (1990, p. 529).

[17] Forgas uses 'affect' as 'a generic label to refer to both moods and emotions' (p. 41).

[18] This experiment is reported in Schwarz and Clore, 1983.

Forgas attempts to demonstrate that 'affect infusion is a significant and reliable source of judgmental distortions,' and his model 'predicts that affect infusion should not influence judgments based on ... motivated processing strategies' (p. 51). The 'specific goals' he cites as motivators of processing are 'mood repair and mood maintenance, self-evaluation maintenance, ego enhancement, achievement motivation, and affiliation' (p. 47; cf. Bower and Forgas, 2000, pp. 130–5, 138). Forgas wants to accommodate cases in which, instead of mood-congruent processing, incongruence is found; and his explanation of such cases is partly motivational (cf. Bower and Forgas, 2000, pp. 135, 154–5). In mood repair, for example, people selectively attend to memories and thoughts that are incongruent with their unpleasant feelings, motivated by a desire to feel better.

The idea that 'goals' such as these are at work in some cases of lay hypothesis testing is easily accommodated by the FTL model. For people experiencing an unpleasant mood or emotion, or a threat to their positive self-image, certain associated errors may be especially costly. For example, in some cases in which people are feeling sad or guilty, the errors of underestimating the quality of their lives or overestimating their responsibility for a harm may be particularly costly. However, this point about the FTL model should not be taken to ground the claim that there is a strict division of labour in lay hypothesis testing between motivation to minimize costly errors and affect infusion. Return to jealous Jed. Owing partly to his jealousy, the most costly error for him may be falsely believing that Jane is faithful, and his processing may be congruent with his jealousy. Similarly, the costliest error for someone who is feeling particularly proud of himself may be falsely believing something that would entail that his pride is unwarranted, and his processing may be congruent with his pride. The question is open whether there is both motivated processing and affect infusion in these scenarios. Forgas apparently commits himself to holding that if motivated processing is at work in them, affect infusion is not. Seemingly, a significant part of what accounts for his taking this view is the difficulty, in such scenarios, of demonstrating empirically that affective states played an infusing role—that, for example, in cases like Jed's, selective attention to and retrieval of thoughts and images congruent with one's jealousy is accounted for at least partly by affect infusion rather than solely by other factors, including 'motivated processing strategies.'

Again, on Forgas's model, 'judgments based on ... motivated processing strategies' are not influenced by affect infusion (p. 51). If

the FTL theory is correct, all lay hypothesis testing involves motivated processing strategies and Forgas's claim about his model, literally interpreted, leaves no room for affect infusion in that sphere. Of course, the FTL theory may be overly ambitious, and Forgas may have been overly restrictive in his statement of what his model predicts. However this may be, a method for testing for the joint influence of motivated processing and affect infusion in biased belief would be useful.

A related issue also merits further investigation. Return again to Jed. He wants it to be true that Jane is not having an affair, and he presumably fears at some point that she is. Eventually, he comes to believe that Jane has been unfaithful. Suppose that Jed's jealousy contributed to that biased belief. Assuming that his jealousy affected his framing of hypotheses, his attention, or the salience of his evidence in a way that contributed to his biased belief that Jane is unfaithful, why didn't it happen instead that his fear, or a constituent desire, affected these things in a way that contributed to his acquiring a belief that she is faithful? Alternatively, why didn't his fear, or a constituent desire, block the relevant potential effects of his jealousy, with the result that the balance of his evidence carried the day?[19]

A proponent of the FTL theory might answer these questions in a way that downplays belief-biasing roles for emotions. In a typical case of romantic jealousy where there are some grounds for suspecting infidelity, the belief that one's romantic partner is having an affair would cause psychological discomfort, but it might also promote one's chances of taking successful steps to save one's relationship. It may be suggested (1) that what one believes is determined by a combination of (a) the strength of one's evidence for and against the proposition that one's partner is having an affair and (b) which error one more strongly desires to avoid and (2) that b is determined by the relative strengths of one's desire to avoid the psychological discomfort of believing that one's partner is having an affair and of one's desire to maintain the relationship. However, this view of things may be too simple. Perhaps distinctively emotional features of jealousy can influence what a jealous person believes in a way that does not depend on desire. Furthermore, even

[19] It may be suggested that Jed's fear issued in fear-congruent processing that meshed with his jealousy-congruent processing. Even if that is so, one wants to understand why the desire-component of his fear—his desire that Jane is not having an affair—did not contribute to motivated processing resulting in a belief that she is innocent of infidelity, or block effective jealousy-congruent processing.

if desire and desire-strength are relevant to what a jealous person comes to believe, that is consistent with his jealousy's having a 'direct' biasing effect on what he believes.

The questions I raised about Jed are difficult ones.[20] Answers that properly inspire confidence will not, I fear, be produced by philosophical speculation. Nor, as far as I know, are such answers available in the empirical literature on emotion: we need to know more than is currently known about the effects of emotions on cognition. These observations are, of course, consistent both with the truth and with the falsity of the direct and solo emotion hypotheses. I am keeping an open mind and trying to be unbiased.[21]

Bibliography

Barnes, A. 1997. *Seeing Through Self-Deception* (Cambridge: Cambridge University Press).

Baron, J. 1988. *Thinking and Deciding* (Cambridge: Cambridge University Press).

Bower, G. and Forgas, J. 2000. 'Affect, Memory, and Social Cognition', In E. Eich, J. Kihlstrom, G. Bower, J. Forgas, and P. Niedenthal, (eds) *Cognition and Emotion* (Oxford: Oxford University Press).

Dalgleish, T. 1997. 'Once More with Feeling: The Role of Emotion in Self-Deception', *Behavioural and Brain Sciences* **20**, 110–11.

Davis, W. 1988. 'A Causal Theory of Experiential Fear', *Canadian Journal of Philosophy* **18**, 459–83.

Derryberry, D. 1988. 'Emotional Influences on Evaluative Judgements:

[20] A proponent of the direct emotion hypothesis may urge that occurrent *aversions* to specific costly mistakes do the work attributed to *desires* to avoid these mistakes in my sketch of the FTL theory, that these aversions play the role attributed to the avoidance desires in determining confidence thresholds. Such a theorist may also contend that occurrent aversions are emotions and argue that even though an aversion to falsely believing that p has a desire to avoid falsely believing that p as a constituent, the work of the aversion in biasing belief is not exhausted by the biasing work of the desire. The claim may be that some distinct, affective feature of the aversion makes a biasing contribution of its own to confidence thresholds. It may also be claimed that the aversion plays additional direct roles—for example, enhancing the salience of evidence for $\sim p$. This is another issue that requires further investigation, including conceptual spadework. Once it is suggested that occurrent aversions are emotions, the suggestion that all desires—or all desires with some felt intensity—are emotions may not be far behind.

[21] Parts of this article derive from Mele, 1987, 1997a, 2000, 2001, and 2003. I am grateful to audiences at the University of Zurich Ethics Center and the University of Manchester for fruitful discussion.

Roles of Arousal, Attention, and Spreading Activation.' *Motivation and Emotion* **12**, 23–55.
de Sousa, R. 1987. *The Rationality of Emotion* (Cambridge: MIT Press).
Forgas, J. 1995. 'Mood and Judgement: The Affect Infusion Model', *Psychological Bulletin* **117**, 39–66.
Friedrich, J. 1993. 'Primary Error Detection and Minimisation (PED-MIN) Strategies in Social Cognition: A Reinterpretation of Confirmation Bias Phenomena', *Psychological Review* **100**, 298–319.
Gigerenzer, G. and Hug, K. 1992. 'Domain-Specific Reasoning: Social Contracts, Cheating, and Perspective Change', *Cognition* **43**, 127–71.
Gur, R. and Sackeim, H. 1979. 'Self-Deception: A Concept in Search of a Phenomenon', *Journal of Personality and Social Psychology* **37**, 147–69.
Johnston, M. 1988. 'Self-Deception and the Nature of Mind', In B. McLaughlin and A. Rorty, (eds) *Perspectives on Self-Deception* (Berkeley: University of California Press).
Kilter, D., Ellsworth, P. and Edwards, K. 1993. 'Beyond Simple Pessimism: Effects of Sadness and Anger on Social Perception', *Journal of Personality and Social Psychology* **64**, 740–52.
Klayman, J. and Ha, Y. 1987. 'Confirmation, Disconfirmation, and Information in Hypothesis-Testing', *Psychological Review* **94**, 211–28.
Kunda, Z. 1999. *Social Cognition* (Cambridge: MIT Press).
— (1990). 'The Case for Motivated Reasoning', *Psychological Bulletin* **108**, 480–98.
Lazar, A. 1999. 'Deceiving Oneself or Self-Deceived? On the Formation of Beliefs "Under the Influence"', *Mind* **108**, 265–90.
Mele, A. 2003. *Motivation and Agency* (New York: Oxford University Press.
— 2001. *Self-Deception Unmasked* (Princeton: Princeton University Press).
— 2000. 'Self-Deception and Emotion', *Consciousness and Emotion* **1**, 115–37
— 1999. 'Twisted Self-Deception', *Philosophical Psychology* **12**, 117–37.
— 1997a. 'Real Self-Deception', *Behavioural and Brain Sciences* **20**, 91–102.
— 1997b. 'Understanding and Explaining Real Self-Deception', *Behavioural and Brain Sciences* 20: 127–34.
— 1987. *Irrationality* (New York: Oxford University Press).
Nisbett, R. and Ross, L. 1980. *Human Inference: Strategies and Shortcomings of Social Judgement* (Englewood Cliffs: Prentice-Hall).
Pears, D. 1984. *Motivated Irrationality* (Oxford: Oxford University Press).
Quattrone, G. and Tversky, A. 1984. 'Causal Versus Diagnostic Contingencies: On Self-Deception and on the Voter's Illusion', *Journal of Personality and Social Psychology* **46**, 237–48.
Schwarz, N. 1990. 'Feelings as Information: Informational and Motivational Functions of Affective States', In E. Higgins and R. Sorrentino, (eds) *Handbook of Motivation and Cognition*, vol. 2. (New York: Guilford Press).

Schwarz, N. and Clore, G. 1983 'Mood, Misattribution and Judgements of Well-Being', *Journal of Personality and Social Psychology* **45**, 513–23.
Sharpsteen, D. and Kirkpatrick, L. 1997. 'Romantic Jealousy and Adult Romantic Attachment', *Journal of Personality and Social Psychology* **72**, 627–40.
Tesser, A., Pilkington, C. and McIntosh, W. 1989. 'Self-Evaluation Maintenance and the Mediational Role of Emotion: The Perception of Friends and Strangers', *Journal of Personality and Social Psychology* **57**, 442–56.
Trivers, R. 1985. *Social Evolution* (Menlo Park, CA: Benjamin/Cummings).
Trope, Y., Gervey, B. and Liberman, N. 1997. 'Wishful Thinking from a Pragmatic Hypothesis-Testing Perspective', In M. Myslobodsky, (ed.) *The Mythomanias: The Nature of Deception and Self-Deception* (Mahwah, NJ: Lawrence Erlbaum).
Trope, Y. and Liberman, A. 1996. 'Social Hypothesis Testing: Cognitive and Motivational Mechanisms', In E. Higgins and A. Kruglanski, (eds.) *Social Psychology: Handbook of Basic Principles* (New York: Guilford Press).

XI. Emotion, Weakness of Will, and the Normative Conception of Agency[1]

KAREN JONES

Empirical work on and common observation of the emotions tells us that our emotions sometimes key us to the presence of real and important reason-giving considerations without necessarily presenting that information to us in a way susceptible of conscious articulation and, sometimes, even despite our consciously held and internally justified judgment that the situation contains no such reasons. In this paper, I want to explore the implications of the fact that emotions show varying degrees of integration with our conscious agency—from none at all to quite substantial—for our understanding of our rationality, and in particular for the traditional assumption that weakness of the will is necessarily irrational.

The paper has two targets: the proximal target is the claim that *in* choosing the incontinent action *rather than* the continent one, the agent necessarily does something irrational;[2] the distal target is the dominant naturalistic conception of how to justify norms of rationality. I'm taking aim at the latter through the former. The naturalist project aims to articulate and defend norms of rationality that are norms for the kind creatures we are—that is, for finite, embodied, social beings, with a specific cognitive architecture,

[1] I would like to thank David Owen, Laura Schroeter, Francois Schroeter, Sigrun Svarvarsdottir, members of the ANU weakness of the will reading group, and audiences at Hobart, Melbourne, Sydney, and the Royal Institute of Philosophy conference on the emotions, Manchester 2001 for helpful discussion of the themes in this paper.

[2] This claim must be distinguished from the stronger claim that it is always *more irrational* for an agent to act against her all-things-considered judgement as to what she should do than act on the basis of it. The claim that continence is always more rational than incontinence is a strong thesis and one not easily defended given that an agent's all-things-considered judgment can be—quite literally— crazy. The target claim is the weaker—and therefore more plausible—thesis that weakness of will always adds an element of irrationality over and above any irrationality that might be involved in the formation of a poor all-things-considered judgment. The stronger and weaker theses are distinguished by David Owen (n.d.) but not by Nomi Arpaly (2000).

functioning in particular environments. On the standard way of developing that project, norms of rationality are not *a priori* knowable and none can be viewed as privileged. Even norms as well-entrenched as norms prohibiting incontinence must await empirical support and recent work on the emotions raises the possibility that they might fail to get it. I argue that reflecting on what you might be tempted to say about emotion, weakness of will, and rationality—if you take a naturalist approach towards these topics—reveals the need for naturalists to pay greater attention to our practical and epistemic *agency* and to norms grounded in our conception of ourselves as rational agents. Nonetheless, I argue, on naturalist grounds, against the view that weakness of will always adds a element of irrationality over and above any irrationality contained in the agent's all-things-considered judgment.

1. Naturalism and norms of rationality

The naturalist project of understanding what it means for the constrained finite creatures we are to be rational has been pursued in greater depth with respect to questions regarding our theoretical rationality than with respect to questions regarding our practical rationality. So let me begin by giving a quick sketch of the naturalist approach to norms of theoretical rationality before switching to issues in practical rationality.

According to the standard naturalist picture of how norms of theoretical rationality are to be justified it is an *empirical* question whether a putative norm counts as a genuine norm and so merits our allegiance or not: does following that norm, given the kind of creatures we are, operating in the kinds of environments in which we operate, advance or hinder our epistemic goals? These epistemic goals, in turn, are to be discerned through an investigation of our epistemic practices, and include, perhaps most centrally, truth-tracking, but also arguably include concern about the system among and significance of the truths that are tracked.[3] If it turns out that, given the features of the environment in which we operate and given our cognitive equipment, a norm is not truth-conducive, then that norm is no genuine norm. Thus, for example, Louise Antony argues

[3] See Antony, 1993, 2000, for exposition and defence of this naturalistic approach to the question of justifying norms. For a naturalistic discussion of epistemic virtues other than truth-tracking, see Goldman, 1986, Chapter 2.

that a norm of objectivity that enjoins us to eliminate bias in our reasoning would be a norm that fails to assist us in truth tracking: cognitive achievements, whether basic such as the ability to learn a language, or sophisticated such as the ability to produce a well-confirmed scientific theory, are made possible for finite creatures like us only on account of our having biases both native and acquired—for example, acquired as the result of participation in a going scientific research programme.[4] If such biases are central to our ability to come to know truths about the external world, then having them cannot be disparaged as irrational. Nor does the story stop with rehabilitating bias. Recent empirical research shows that 'we cannot get by epistemically without shortcuts and tricks of all kinds, many of which would not survive scrutiny by traditional epistemological lights' (Antony, 2000, 115).[5]

A naturalist inquiry into the preconditions of our cognitive abilities can thus lead to a reconception of what it takes for creatures like us to have and display theoretical rationality and there's no *in-principle* reason to think that the deliverances of this inquiry have to coincide with the deliverances of armchair inquiry into norms of theoretical rationality (and the story about bias suggests they won't). Similarly, naturalist inspired reflection on what we know about how the emotions contribute to the practical rationality of finite creatures like us might lead us to reconceive norms of practical rationality—or so it would seem.

A parallel naturalistic story about practical rationality begins by identifying the goals of practical decision-making and then offers an instrumental justification of norms of practical rationality according as they help or hinder us—again given the kinds of agents we are, operating in the kinds of environments in which we operate—in achieving those goals. There's no non-controversial and adequately determinate account of the goals of practical decision-making. But even a statement of the goals that is schematic and indeterminate enough to secure consensus across controversy, is enough to set up the problem.[6]

[4] Antony, 1993, 2000. For a discussion of the theory dependence of method and the need for presuppositions, see Boyd, 1983. An overview of these issues with selected further reading is found in Boyd, Gasper and Trout, 1991.

[5] The literature here is long, but see especially Gigerenzer, *et al.*, 1999; Stein, 1996.

[6] Thus, for the purposes of this argument, I need not take a stance on whether there are external reasons or whether all reasons are internal (Williams, 1980).

We can make progress understanding the goals of practical decision-making by considering what it is that agents regret about their decisions, for regret indicates failure to do what one was trying to do. When an agent faces a practical problem, she is trying to *identify* and select an action option that responds to *all* the reason-giving considerations present in the situation, in proportion to their strength as reasons. In hard choices, there may be no action option that respects all the reason-giving considerations and the agent must select that action option that answers to the most weighty of her reasons.

When we face a practical problem, we are not just trying to act for what we *believe* to be good reasons: we are trying to act for what actually are good reasons for us (but of course the central, though not, I'll argue, the only path of access to these reasons is through our beliefs about them.) Nor are the considerations that can count as reasons limited to those that mesh with or answer to concerns that the agent *currently* values. Agents regret overlooking considerations that answer to currently valued concerns ('I just didn't see how my doing that could help you, I'm so sorry'), but equally they regret recognizing considerations as reason-giving on the basis of concerns that fail to survive reflective scrutiny ('I can't believe I was so worried about what he might think of me') and they regret choosing in ignorance of values ('I didn't appreciate the importance of family until it was too late'). As epistemic agents we are trying to latch onto truths about the world, though all we have to go on in achieving this goal are our own mechanisms and methods, reliable and otherwise, for detecting such truths, together with our own best take on the limits and liabilities of our methods and mechanisms. Likewise, as practical agents, we are trying to latch onto those considerations that really are reason-giving for us in a situation, yet all we have to go on in achieving this goal is nothing more than our own mechanisms and methods, reliable and otherwise, for detecting these considerations together with our own best take on the limits and liabilities of the methods we use to work out what considerations matter.[7] The account remains indeterminate and so minimalist because it leaves it open how to cash out the phrase 'really are reasons for an agent.'

[7] This point is not meant to imply epistemic individualism: an important method for acquiring knowledge, even knowledge about practical matters, is testimony and the task of working out the reliability of these mechanisms and methods is conducted socially, and rests on divisions of cognitive labour.

Now, combine this minimalist account of what we are trying to do in practical decision-making with the naturalist story of how to defend norms of rationality and you get the result that it is an *empirical* question whether we are such that our all-things-considered judgments reliably help or hinder us in latching onto our reasons and thus whether those judgments should be accorded normative authority. Finally, add some observations and examples, both scientific and everyday, about the ways emotions contribute to our being able to track our reasons, not despite, but *because of* the fact that they display at most partial integration with our evaluative judgment, and one might be led to the conclusion that our all-things-considered judgment has no special normative standing and thus that acting against it is not necessarily irrational. Here are the observations and examples:

(1) Emotions exhibit varying degrees of integration with our conscious deliberative faculties and sometimes their very independence from those faculties contributes to their adaptiveness. For example, fear responses can be initiated before the stimulus is processed by the visual or auditory cortexes and the speed of response enabled by this feature of our hardwiring is unquestionably adaptive. Even brain damaged patients who are unable to form long-term memories can have functioning fear systems that enable affective learning that 'tracks' their practical reasons though without generating any higher-level understanding of that tracking. For example, Joseph LeDoux (1996) reports the case of a woman who though unable to recognize her doctors from one meeting to the next, was able to learn not to shake hands with a doctor who had previously pricked her with a tack concealed in his palm.

(2) Even emotions that cannot be had without considerable cognitive sophistication, such as resentment and indignation which require the agent to construe the situation in terms of relatively sophisticated evaluative concepts, nonetheless display only partial integration with the agent's conscious evaluative judgment. We can be resentful of persons who we sincerely judge have done nothing that merits resentment; indeed, it has been reflection on these kinds of cases, as well as cases involving phobic emotions, that has driven the current movement in philosophy of the emotions away from judgmentalist accounts that analyse the cognitive content of an emotion in terms of evaluative belief.[8] Moreover, this ability

of even sophisticated emotions to run in opposition to evaluative judgment turns out to be important for our practical rationality. Such "outlaw" (Jaggar, 1996) emotions can function as correctives to internally coherent but false and often ideologically driven views about, to mention just two examples explored in the philosophical literature, the status of women (Scheman, 1980) and of slaves (Bennett, 1974). Our efforts to make sense of outlaw emotions can provide a starting point for the critical re-examination of even quite central evaluative assumptions. Thus, emotions can function as recalcitrant data that force a change in our evaluative assumptions.

(3) In particular cases, an agent's emotions can be keyed-to her reasons in such a way that they enable the agent to track those reasons, while her all-things-considered judgment does not. Nomi Arpaly gives us the example of Emily, who has always believed that she should pursue a PhD in chemistry. Once embarked on this project, however, Emily finds herself feeling 'restless, sad, and ill-motivated' (Arpaly, 2000, 504) to stick with her studies. As Arpaly describes the case we are to suppose that Emily's affective discomfort is keying her to the reasons that she has to leave the program—her feelings are responses to the fact that the program is ill-suited to her talents, preferences, and character. Yet this evidence does not secure uptake in her judgment. She herself sees her feelings as 'groundless.' Emily acts on her feelings and leaves the program. Later she comes to understand the reasons for her feelings and 'cites them as the reasons for her quitting, and regards as irrationality not her quitting but that she held on to her conviction that the program was right for her for as long as she did.' (Arpaly, 2000, 504).

Nor are cases of this kind uncommon: often our gut-feelings key us to the presence of reasons even though we cannot, at the time,

[8] If we can be resentful without the conscious belief that the person has done anything to merit resentment, and if, as judgmentalists suppose, resentment is constituted in full in or in part, by an evaluative belief, then we will have to suppose that the belief in question is unconscious. But there are good reasons for not attributing such unconscious beliefs in all cases and only the very judgmentalist theory that is in dispute for supposing that there would have to be such beliefs. For an argument against attributing unconscious beliefs as promiscuously as this theory would require, see Greenspan, 1988, chapter 2. For discussion of emotions in opposition to evaluative judgment, see Calhoun, 1984, and Stocker, 1996.

articulate what those reasons are, and even though our conscious deliberative judgment tells us no such reasons obtain (you're just being silly, get a grip). Sometimes we act on these feelings against our better judgment and discover to our relief that our feelings were exactly right and that our weakness saved us from acting on the basis of a misguided all-things-considered judgment. That is, often enough our emotions and not our conscious deliberative judgment are what enables us to latch onto those reason-giving considerations that we ought to recognize. Thus, if we are committed to the naturalist project of defending norms of rationality instrumentally according as how they help or hinder us in achieving the goals of practical deliberation, then there might be real grounds for thinking that—like norms ruling out bias—norms that prohibit incontinence might not be the sort of norms that enable the kind of limited, finite, embodied agents that we are reliably to latch onto our reason-giving considerations. What normative status to accord our all-things-considered judgment becomes an empirical question and we might answer that it has no particular normative standing.

Thus, thinking about the role of emotions in keying us to our reasons might generate an account of rationality with the following features:

(i) *no necessary irrationality*: in choosing the incontinent action over the continent one, the agent does not necessarily display irrationality, as the incontinent action may be produced by a well-functioning mechanism that is reliably keying the agent to her reasons, when her all-things-considered judgment is not.

(ii) *no transparency or agential privilege*: the rationality or irrationality of an action can be very hard to discern since it will depend on whether the mechanisms that lead to the action are keying her to reasons or not. The agent may be in the *worst* position to determine this: if she chooses the incontinent action she will think she is being irrational (as Emily did) but she might be quite wrong about this.

(iii) *broad supervenience base*: the rationality of an action supervenes on a comparatively broad class of facts including most especially facts about the reliability of the mechanism that produced the action and facts about how the agent formulated her all-things-considered judgment.

(iv) *no special normative standing for all-things-considered judgment*:

> a theory of rationality should not assume that there is something special about an agent's best [all-things-considered]

> judgment. An agent's best judgment is just another belief, and for something to conflict with one's best judgment is nothing more dramatic than ordinary inconsistency between beliefs, or between beliefs and desires (Arpaly, 2000, 512)[9]

2. A problem with this picture: the normative conception of agency

I think that we should be quite worried about the picture that has emerged: to suppose, as Arpaly does, that well-functioning mechanisms capable of latching on to reasons can be sufficient for rationality and that our all-things considered judgments are normatively speaking on all fours with our other beliefs, or to suppose that whether they are on all fours normatively is an empirical question that requires further investigation as the naturalist account that's so far been on the table does, is not yet to recognize our epistemic and practical *agency*.[10]

As a reflective agent, I cannot view myself as merely a system—however well functioning—of sub-systems, that passively register and respond to environmental stimuli much as a thermostat

[9] Arpaly would assent to each of 1–4, though she is operating within a different framework from the one used here. At least for the purposes of argument, Arpaly assumes that an act is rational if it maximizes the satisfaction of the agent's desires and she takes the chief argument against the rationality of incontinence to be an argument from (in)coherence. But this is too narrow a conception of practical rationality and the main argument against incontinence is not, as we shall see, best formulated in terms of coherence. A full exploration of Arpaly's argument would take me too far from the main business of the paper to be undertaken here.

[10] For a response to my pressing this objection to the naturalist epistemological project, see Antony, 2000. My current formulation of the problem has been influenced by Louise Antony's explication of it. However, Antony puts the point in terms of 'transparency': 'Commitment to rationality involves, among other things, a norm that bids us make our reasons transparent to ourselves as we reason—arguably that is what reasoning *is*.' (114–15) I think that the point is better put in terms of a commitment to rational guidance by reasons seen as such. Some failures of transparency will be failures to guide action (belief) by reasons seen as reasons; e.g. when I'm moved by psychological forces that are mysterious and opaque to me. But framing the issues under the concept of transparency raises further issues (e.g. about the extent to which we can know about the machinery that subserves our rational processes) that seem to me orthogonal to the central issue of rational guidance.

registers and responds to changes in temperature. Nor can I view my reasoning self—that part of me that engages in conscious deliberation about what to do or what to believe—as simply one additional epistemic mechanism operating side-by-side with other mechanisms such as perceptual or emotional ones. To think of myself in this way is not to think of myself as an agent at all. It is to give up thinking of myself as rationally *guiding* my actions via reasons. Yet from the first person point of view, it seems to me that my conscious deliberative self is capable of guiding my action; moreover, it seems that I am capable of guiding my action in accordance with my best reasons, reasons not merely registered, but understood as reasons, that is, understood as *justifying* the performance of an action. Let's see if the thought can be made more precise.

Distinguish two kinds of agents differentiated by the relation in which they stand to reasons: the first kind of agent guides its action via a conception of its reasons as reasons. Agents of this kind must have the capacity for reflection, for *guiding* actions via reasons *seen as reasons* requires that the agent have a self-conception and posses the concept of a reason as something that justifies the performance of an action. Given the comparative nature of such justifications, guiding actions via reasons understood as conferring justification on the performance of one action rather than another, commits an agent to guiding actions via best reasons.[11] That is, if an agent is to have a justification for doing A *rather than* B or C, she must suppose that her best reasons support A, even though there may be something to be said for (some reasons in favour of) both B and C.

Call agents who guide their actions via reasons understood as reasons, *reason-responders*. Reason-responders must possess and exercise a complex set of capacities if they are to respond to reasons understood as reasons. Among these capacities are the capacities to step back from any actional impulse and inquire whether the desire really reflects anything choiceworthy in the action (e.g. is the desire

[11] Talk of 'best reasons' here is consistent with but by no means requires a maximizing conception of rationality according to which rational agents seek to maximize some single value which renders all values commensurable, such as happiness, or utility. I think such views make a mistake about the nature of value. My point is less controversial: that if one is to have a reason to do A *rather than* B, and one has a reason to do both (e.g. doing A would be fun, doing B would help make an important deadline) then if one is to choose A rather than B that must be because 'having fun' is in this context seen as a better reason than all its competitor reasons; that is, is seen as the best reason.

to be eliminated rather than satisfied, as are desires to smash opponents in the face with tennis rackets (Watson, 1975)). Further, the agent must be sensitive to when putative reasons are defeated and when they are outweighed. For example 'it looks green' is defeated as a reason for believing it is green the experimenter has just told me the contact lenses in my eyes will give everything a greenish hue. Thus, having sensitivity to when reasons are defeated and when they are outweighed requires the capacity to reflect on the status of the deliverances of those mechanisms that purport to latch onto reasons such as perception, emotion and desire, but also the capacity to reflect on reasoning itself—for it too can deliver false representations of the reasons that obtain. In sum, to be able to respond to reasons as reasons, an agent requires critical reflective ability, dispositions to bring that ability to bear when needed, and dispositions to have the results of such reflection control their behaviour.[12]

The second kind of agent is capable of registering reasons and behaving in accordance with them, but it need posses neither the concept of a reason nor have a self-conception. It thus need not have the higher-order reflective capacities characteristic of reason-responders. Call such agents *reason-trackers*.

Reason-responders are thus highly sophisticated reason-trackers; that is, agents capable of tracking reasons in virtue of responding to them as reasons. The advantage of being a reason-responder rather than a mere tracker is that responders will display *robustness* in their ability to track their reasons. Animals, for example can reason track through innate and learned behaviour, but a creature who lacks critical reflective capacities will not be able to display the same kind of flexibility in its action and sensitivity to the implications of changes in its environment that a reason-responder can.

With the distinction between reason-responders and reason-trackers on hand, we can now return to the task of making the thoughts about the nature of our agency, as it appears to us from the inside, at least a bit more precise. From the first person point of view, I conceive of myself as capable of being a reason-responder; that is, I *can* guide my action in accordance with reasons understood as such. Moreover, insofar as I take myself to be rational, I take

[12] A number of theorists are moving towards something like the distinction between what I'm calling reason-responders and simple animal agents, who in my terms are merely 'reason-trackers.' See Tyler Burge, 1996; Christine Korsgaard, 1996, 1997; Thomas Scanlon, 1998, Chapter 1, especially at 23; Joseph Raz, 1999; and Francois Schroeter, n.d., who offers the most extended discussion of the capacities required to act for reasons seen as such.

myself to *be* a reason-responder. Thus, any story about my rationality that is only story about whether and how my subsystems reason-track is inadequate as a story about the kind of agent I conceive of myself as being. But, I cannot conceive of myself as a reason-responder without being committed to guiding my actions via my best reasons, for to the extent that I fail to live up to this commitment, I fail to be the *kind* of agent that I take myself to be. This commitment thus follows from my conceiving of myself as a reason-responder. In this way, it seems that not all norms of rationality are a posteriori and await empirical proof of their usefulness: in virtue of conceiving of myself as a reason-responder I am committed to a norm that enjoins me to guide my actions via my best reasons and that says that when I fail to do so I am not rational. I can see some norms—whether norms of practical or of theoretical rationality such as the norms about bias discussed earlier—as norms that are going to help (or hinder me) in responding to my best reasons. And I can lose my allegiance to them once I see how they hinder me from latching onto my reasons, but I can't give up on this norm of rational guidance without giving up on my conception of myself as having the kind of agency I take myself to have.

It is important that this argument is phrased in terms of how I *conceive* of myself. There's a real question, and it is not immediately answerable first personally, whether I actually *am* a reason-responder. Perhaps my conscious deliberation has got nothing to do with what this body that is mine subsequently does; perhaps the appearance of guidance, is just that, *appearance*. Indeed, that's what is so disquieting about empirical research showing we are skilled confabulators and about research into the determinants of our action, such as research which purports to show that, unless you select the box of cereal on the special display stand, then you will almost certainly choose the cereal on the top shelf at the end of the aisle. (When I read about that research it so happened that the cereal I buy was to be found on the top shelf at the end of the aisle, though I would have said I was buying it because it is high in fibre, and low in sugar and fat, and so it is.) Thus, there is a real question how to reconcile this normative conception of my own agency, which seems first-personally given, with third-personal accounts of the determinants of action. Moreover, by conceiving of myself as a reason-responder, and so being committed to guiding my action via best reasons, there's a chance that I might fail to be a reason-tracker. Perhaps I'd do better if I stopped attempting to guide my actions via reasons and just let my well-functioning reason-tracking mechanisms take over. I'll return to the question of

whether we should continue to conceive of ourselves as reason-responders later in the paper.

3. Explicating the core commitment

I have claimed that if an agent is to conceive of herself as a reason-responder, then she must be committed to guiding her action via best reasons. But on one natural reading of what this commitment amounts to, it is clearly violated in cases of weakness of will: for the commitment to guide one's action via best reasons is readily understood as the commitment to guide one's action via one's all-things-considered or *best judgment* as to what one has reason to do. No doubt, it is this conception of rational guidance that explains why incontinence has been assumed to be *so* manifestly irrational that no argument for its irrationality need be given (though it is not among the many reasons Arpaly herself canvasses). Call this interpretation of the commitment, the intellectualist reading.[13] On the intellectualist reading, any agent who acts against her best judgment thereby fails to live up to this constitutive commitment. However poorly formed her all-things-considered judgment, in failing to act according to it, she adds a further failure of rationality to her failure in forming the ill-advised judgment. Moreover she adds a *failure of an especially serious kind*: she fails in respect of the very norm that defines what it is to be the kind of agent she takes herself to be. (So most often, then, it will be all-things-considered more irrational to act incontinently than continently, but—see note 1—this need not be so in every case.)

The intellectualist reading of this core commitment denies each of the four claims about rationality mentioned earlier, asserting instead:

(i) *necessary irrationality*: in choosing the incontinent action over the continent one, the agent necessarily displays irrationality; indeed, irrationality of an especially serious sort.
(ii) *transparency and agential privilege*: rationality or irrationality—insofar as these concern failure or success at guiding one's action via one's best judgment—is readily discernible and inasmuch as the agent has privileged access to her all-things-considered judgment, she is in the best position to know the rational status of her action.[14]
(iii) *narrow supervenience base*: the irrationality of an action (in respect of this central norm) supervenes on a relatively narrow range of facts: what was her best judgment? What was her action? Do they match?
(iv) *privileged normative standing for all-things-considered judgment*: all-things-considered normative judgments get normative authority just in virtue of the kind of judgments they are.

[13] The intellectualist position is common: it is explicitly endorsed by Korsgaard (1997, 222); Raz, 2000, 16; Scanlon, 1998, esp. 25 and is assumed by Wallace, 1999.

On the intellectualist reading, there's a single strand to the commitment to guiding one's actions via best reasons; namely the disposition to have one's all-thing-considered judgment be *authoritative* with respect to what one subsequently does. The all-things-considered judgment is seen as having normative authority for the agent *just in virtue* of its being the deliverance of the agent's conscious reasoning self. The commitment to rational guidance is thus seen as having unique expression in acts of continence (or most perfect expression: there is a weaker position available here, but I leave it to one side as the argument goes through even against this slightly more complex position). But, to use a rather uncharitable analogy, this is much like saying that one's concern for another finds unique (or most perfect) expression in what one says. Notoriously, this is not the case.

I want to outline a third picture, distinct from both of the currently available accounts, for it seems that there should be room for an intermediate position which, like the intellectualist position, recognizes the importance of our commitment to rational guidance via reasons but which has a richer understanding of what that guidance amounts to. The alternative picture shares with the simple naturalist model the following claims:

(i) no necessary irrationality,
(ii) no transparency or agential privilege, and
(iii) broad supervenience base.

[14] The qualification, 'insofar as these concern failure or success at guiding one's action via one's best judgement' matters here (and in (iii) below). The intellectualist does not think that norms of rational guidance are the only norms of rationality, and there may be failures of transparency and privilege with respect to success at other norms. However, the intellectualist position is typically combined with epistemological internalism and thus is typically combined with the view that the agent has in principle access to the rational status of her actions and beliefs.

But, like the intellectualist account, it recognizes the centrality of the commitment to rational guidance, though it finds expression of that commitment in more activities than the intellectualist recognizes as expressive of it.

On the alternative account of rational guidance that I propose, the commitment to rational guidance is to be understood as the commitment to the on-going cultivation and exercise of habits of reflective self-monitoring of our practical and epistemic agency. That is, the commitment to rational guidance is the commitment to the on-going cultivation and exercise of whatever abilities it is that enable the agent to have and display the capacities that are characteristic of reason-responders (see section 2). It is an empirical matter what dispositions will enable agents *of the kind we are* successfully to reflect on the status of the deliverances of those mechanisms that purport to latch onto reasons—including reasoning itself. For example, there's evidence to suggest that our ability to be wise in our trust of our own judgment and of the judgment of others rests on emotional capacities such as the capacity for empathy; thus, commitment to rational guidance includes commitment to the cultivation and exercise of empathy. In this way, the dispositions that constitute the commitment to rational guidance will be many and various and they won't all be dispositions of intellect.

If the commitment is multi-stranded in the way I am suggesting it is, then it can find expression in more ways than through the agent guiding her action via her best judgment; indeed, such judgment can *fail* to express the commitment to rational guidance. On the proposed model, an all-things-considered judgment does not get normative authority for free. It has to earn such authority and it earns it in virtue of being the product of a conscious reasoning self that has itself been subject to regulation by reflective capacities.

The difference between the three models can be shown by considering what each model has to say about fast or habitual action; that is, action undertaken without deliberation, either because the situation is urgent and so deliberation is impossible, or because the situation is routine and so deliberation is unnecessary.

On the intellectualist model fast or habitual action expresses the agent's commitment to guiding her action via reasons just in case such action is the result of a reflectively endorsed policy. If I have a reflectively endorsed, for example, acting out of immediate responses of concern for my children in such and such circumstances, then action that results from those motives, even in the absence of formulating an all-things-considered judgment as to what to do nonetheless expresses my commitment to guidance via

best judgment.[15] A similar story can be told about fast action: for example, I might resolve immediately to go with my gut feelings of suspicion in my work as a security guard *because*, having reflected on my track record as a detector of possibly suspicious behaviour and on the costs and consequences of deliberating under these circumstances, I judge that a context-specific policy authorizing immediate action from this affective motive is the best policy. Thus, for fast or habitual action to express an agent's commitment to rational guidance on the intellectualist account it must be the product of self-conscious policies that are themselves endorsed by judgment.

This intellectualist account of the rationality of fast or habitual action contrasts with the account that might be given by someone who thinks that well-functioning reason-tracking mechanisms are sufficient for rationality; not surprisingly, it contrasts with the account explicitly endorsed by Nomi Arpaly. According to Arpaly, no reflective policy is needed: the action is rational if produced by a reliable mechanism and that's all there is to be said. No commitment to guiding actions via reasons is recognized and thus the question of whether such action expresses that commitment does not arise. But this seems to give us the wrong answer: an agent might have a well-functioning reason-tracking mechanism and yet it not be *responsible* for her to take the deliverances of that mechanism to be reason-tracking. It's deliverances would be undermined (the parallel with belief undermining is intended). This could be the case with our security guard: she might have evidence that her suspicion fails to track and if this is so, then her continuing to act on the basis of the deliverances of her emotional sensitivities would be an irresponsible failure of rational guidance. If her self-monitoring dispositions were functioning as they should, then she would cease to trust her emotional sensitivities.

On my preferred third picture, fast and habitual action can express the commitment to rational guidance and will do so just in case the agent's dispositions to reflective self-monitoring are such that she would not rely on that first order sub-system were it reasonable for her to believe that it failed to reason-track. That is,

[15] The intellectualist need not say that agents must express their commitment to rational guidance in all domains and can allow that sometimes one should just be spontaneous (and need not have a policy about just when to be so). However, my argument does not rest on saddling the intellectualist with the further view that all action should express this commitment—the dispute is over what can express it, rather than over the domain in which it should be expressed.

her conscious reflective capacities exert *regulative guidance* over the first-order mechanism, stepping in when necessary to discount those mechanisms, and where possible, to recalibrate them into reason-tracking mechanisms through habituation. Sometimes this guidance may remain 'virtual'—that is, revealed in how the agent would behave in various counter-factual circumstances (were she to have evidence that they are unreliable, for example, but evidence she never gets since they are reliable). There's nothing mysterious about this kind of guidance: it is just one way of describing what happens when an agent's emotional responses are shaped, fine-tuned, and sometimes even radically transformed through the process of character formation so that they become reliable at latching on to the reasons that obtain for her. But we can see at once how action that results from such regulated first-order mechanisms can express the agent's commitment to rational guidance via reasons: our subsystems can reason-track because we, as agents, reason-respond.[16]

The preferred third model has implications for what we should say about the rationality of some cases of incontinence. Regulated sub-systems that reason-track because we reason-respond can be no less operative in generating action when there is an all-things-considered judgment that opposes the action so produced. That is, the functioning of such sub-systems does not stop being expressive of our commitment to rational guidance just because there is now an opposing all-things-considered judgment. In some cases that all-things-considered judgment may be such that the agent would distrust it, if her self-monitoring capacities were functioning as they should. Thus, the regulated sub-system can be more expressive of the agent's commitment to rational guidance than the all-things-considered judgment: the incontinent action can display the agent's commitment to rational guidance more fully than does the continent action.

We can generate a schema for producing examples of rational incontinence: incontinence will be rational just in case: (1) the action is produced by a sub-system that reason-tracks because the agent reason-responded, and (2) the agent would have distrusted her all-things-considered judgment were her self-monitoring

[16] To be precise, but at the cost of the elegance of the slogan, the 'because' should sometimes be read as an initiating because, and sometimes as a maintaining one. That is, sometimes we, as agents, initiate a method, or recalibrate a mechanism in order to latch-on to our reasons; other times, a mechanism will be maintained in place under the 'virtual' guidance of our reflective self-monitoring capacities.

dispositions operating as they should. Once you see how to construct such examples, you can find them all over the place—for example, feminist anger can undo a decision not to raise a certain topic at a meeting and subsequent reflection on what happened can reveal that the decision not to raise the issue was the product of cowardice rather than of a sober assessment of the merits of investing scarce credibility resources to fight this fight rather than some other one. Of course, it won't be easy for the agent to work out whether her action is rational and she will have no privileged access to its rationality—but that result seems to me unsurprising. Further, whether the action is rational supervenes on a complex set of facts about the dispositions that were operative in generating the action. Often these dispositions will reveal themselves in what happens *next*: any agent committed to guiding action via reasons will experience disquiet at her own incontinence. She will want to reflect on that action —and such reflection may indeed reveal that the action was irrational. I am certainly not claiming that most incontinent action is rational. Most of it is not. What I am claiming is only that such action does not necessarily fall afoul of what seems from the first person point of view to be a non-negotiable commitment, a commitment that follows from conceiving of ourselves as having a certain kind of epistemic and practical agency.

If these reflections are along the right lines, then there's another strategy open to anyone who would deny the commonly made assertion that emotions are, if not outright irrational in themselves, then frequent contributing causes of irrationality insofar as they are frequent contributing causes of incontinence. It is salutary to remind anyone who charges emotions with causing incontinence that we frequently fail to do what we judge we ought to do because we cannot summon the compassion or the anger required to do it. Thus emotions can help us to act on our all-things-considered judgment as well as hindering us from doing so. But if I'm right, then we can *also* say: sure, emotions sometimes contribute to incontinence, but that may be just what we need to get us to overcome poor all-things-considered judgment and cannot be assumed necessarily to be irrational.

The argument I've presented is conditional: if we are to conceive of ourselves as reason-responders then we must be committed to the norm of guiding action via best reasons. First personally, it seems that we do conceive of ourselves as agents of this kind, at least insofar as we think of ourselves as rational. I've argued that this self-conception gives rise to norms that do not await further instrumental justification through a demonstration of their

usefulness (as all norms must on one standard naturalist account of them): you get the norm in virtue of the self-conception. But this pushes the question back one step further. Should we think of ourselves as being reason-responders, given that thinking of ourselves in this way brings with it a commitment to the cultivation and exercise of habits of reflective self-monitoring?

There's an instrumental argument to be given here: if you think of yourself this way, then, unless your reflective capacities are really deficient or unless you are unfortunate enough to inhabit some kind of demon-world in which non-reflective sub-systems reason-track while the demon makes sure you'll mess things up if you try and reflect on their reliability, the chances are you will be more nearly able to reason-track than you would if you did not. Attempting to reason-respond can bring it about that you are better able to latch-on to your reasons. And it can bring the benefits of robustness to your reason-tracking abilities. I think the instrumental argument is fine, as far as it goes: we *are* better off thinking of ourselves in this way. I'm willing to bet the farm we're not so stupid or so unlucky that this commitment is a liability. But there is more to be said here—and I think that saying it is compatible with a naturalist approach to normativity. That more is this: in affirming the value of the normative commitment to guiding action (or belief) via best reasons, we affirm the value of the kind of agency we take ourselves to have. We do not *have* to affirm the value of this kind of agency, but if we fail to affirm it, then we fail to affirm the value of something valuable. And that's a non-instrumental reason for affirming the value of this central norm.[17]

Bibliography

Antony, Louise. 1993. 'Quine as Feminist: The Radical Import of Naturalized Epistemology', In Louise Antony and Charlotte Witt (eds) *A Mind of One's Own* (Bolder, Co: Westview Press).

[17] This answer belongs in a family of answers first sketched by Louise Antony, as follows: 'Since it is the active and self-conscious consideration of reasons that makes one an epistemic agent, the norms of rationality can be said to express our conception of what it is to be an epistemic agent; and to endorse that norm is to express one's commitment to the value of such agency.' (2000, 127). Antony's formulation is neutral between a realist gloss on evaluative judgments and an expressivist one. For reasons that cannot be explored here, I have offered a self-consciously realist statement of the evaluative mistake we make if we fail to affirm the value of this kind of agency.

Antony, Louise. 2000. 'Naturalised Epistemology, Morality, and the Real World', *Canadian Journal of Philosophy*, Supplementary Volume 26, 103–37.

Arpaly, Nomi. 2000. 'On Acting Rationally Against One's Best Judgement', *Ethics* **110**, 488–513.

Bennett, Jonathan. 1974. 'The Conscience of Huckleberry Finn', *Philosophy* **49**, 123–34.

Boyd, Richard. 1983. 'On the current status of the issue of scientific realism', *Erkenntnis* **19**, 45–90.

Boyd, R. Gasper, P. and Trout, J. D. (eds). 1991. *The Philosophy of Science* (MIT Press: Cambridge).

Burge, Tyler. 1996. 'Our Entitlement to Self-Knowledge', *Proceedings of the Aristotelian Society* **96**, 91–116.

Calhoun, Cheshire. 1984. 'Cognitive Emotions?' In Calhoun and Solomon 1984, 327–42.

Calhoun, Cheshire. 1989. 'Subjectivity and Emotion', *The Philosophical Forum*, Vol. 10, 195–210.

Calhoun, Cheshire, and Robert Solomon, (eds), 1984. *What is an Emotion? Classic Readings in Philosophical Psychology* (New York: Oxford University Press).

Descartes, Rene. 1985[1649]. *The Passions of the Soul*. In John Cottingham, Robert Stoothoff and Dugald Murdoch (trans.) *The Philosophical Writings of Descartes* Vol. 1. (Cambridge: Cambridge University Press).

de Sousa, Ronald. 1979. 'The Rationality of Emotions', In Rorty 1980, 127–52.

de Sousa, Ronald. 1987. *The Rationality of Emotion* (Cambridge: MIT Press).

Elster, Jon. 1999. *Alchemies of the Mind: Rationality and the Emotions* (Cambridge University Press).

Frank, Robert. 1988. *Passions within Reason* (New York: Norton).

Gigerenzer, Gerd, Todd, Peter and the ABC Research Group. 1999. *Simple Heuristics that Make us Smart* (New York: Oxford University Press).

Goldman, Alvin. 1986. *Epistemology and Cognition* (Cambridge: Harvard University Press).

Greenspan, Patricia S. 1988. *Emotions and Reasons: An Inquiry into Emotional Justification* (New York: Routledge).

Jaggar, Allison. 1996. 'Love and Knowledge: Emotion in Feminist Epistemology', In Ann Garry and Marilyn Pearsall (eds), *Women, Knowledge, and Reality*. Second Edition. (New York: Routledge), 166–190.

Korsgaard, Christine. 1996. *The Sources of Normativity* (Cambridge University Press).

Korsgaard, Christine. 1997. 'The Normativity of Instrumental Reason', In Garret Cullity and Berys Gaut (eds) *Ethics and Practical Reason*. (Oxford: Claredon Press).

Le Doux, Joseph. 1996. *The Emotional Brain: The Mysterious Underpinnings of Emotional Life* (New York: Simon and Schuster).

Moran, Richard. 1988. 'Making Up Your Mind: Self-Interpretation and Self-Constitution', *Ratio*, n.s. 1, 135–51.
Owen, David. n.d. 'The Authority of Practical Judgement'.
Raz, Joseph. 2000. 'When we are ourselves: the active and the passive', In his, *Engaging Reason* (Oxford: OUP), 5–21.
Scanlon, T. M. 1998. *What We Owe to Each Other* (Cambridge, MA: Harvard University Press).
Scheman, Naomi. 1980. 'Anger and the politics of naming', In Sally McConell-Ginet, Ruth Borker and Nellie Furman (eds) *Women and Language in Literature and Society* (New York: Praeger), 174–87.
Schroeter, Francois. 2001. 'Normative Concepts and Motivation'.
Solomon, Robert C. 1973. 'Emotions and Choice', In Rorty 1980, 251–81.
Solomon, Robert C. 1976. *The Passions* (Garden City, New York: Anchor/Doubleday Press).
Stein, Edward. 1996. *Without Good Reason* (Oxford: The Claredon Press).
Stocker, Michael with Elizabeth Hegeman. 1996. *Valuing Emotions* (Cambridge University Press).
Wallace, R. Jay 1999. 'Three Conceptions of Rational Agency', *Ethical Theory and Moral Practice* **2**, 217–42.
Watson, Gary. 1975. 'Free Agency." *Journal of Philosophy* **72**, 205–22.
Williams, Bernard. 1980. 'Internal and External Reasons', In his *Moral Luck: Philosophical Papers 1973–1980*, 101–13. (Cambridge: Cambridge University Press), 1981.

XII. Narrative and Perspective; Values and Appropriate Emotions

PETER GOLDIE

> To the realists.—You sober people who feel well armed against passion and fantasies and would like to turn your emptiness into a matter of pride and ornament: you call yourselves realists and hint that the world really is the way it appears to you. As if reality stood unveiled before you only, and you yourselves were perhaps the best part of it ... But in your unveiled state are not even you still very passionate and dark creatures compared to fish, and still far too similar to an artist in love? And what is 'reality' for an artist in love? You are still burdened with those estimates of things that have their origin in the passions and loves of former centuries. Your sobriety still contains a secret and inextinguishable drunkenness. Your love of 'reality', for example—oh, that is a primeval 'love' ... Subtract the phantasm and every human contribution from it, my sober friends! If you can! If you can forget your descent, your past, your training—all of your humanity and animality. (F. Nietzsche, *The Gay Science*, Book Two, extract from Section 57)

I

We are reflective creatures, capable of thoughts about thoughts, feelings about feelings, and emotions about emotions. Of course, we can be unreflectively engaged in daily interaction with the world, and most of us often are. But our capacity for reflection gives rise to something of a need: a need to understand our lives though reflection on what has happened. So we can agree both with Kafka when he said that our daily life is the only life we have, and also with Kierkegaard when he said that we live our lives forward, but understand them only backwards. We find an extreme case in Leontes, who, in Act III of *The Winter's Tale*, was only able to understand his jealous rage for what it was after it was over; only then could he say 'I have too much believed my own suspicion'. But, by choosing this example, I do not intend to encourage the idea that I mean the domain of emotions to include just those short-term episodes of