'The Logic of Excuses and the Rationality of Emotions'

by John Gardner
Professor of Jurisprudence
University of Oxford
http://users.ox.ac.uk/~lawf0081

Journal of Value Inquiry 43 (2009), 315

I was unhappy with final edits to the published version, which were demanded by the journal's house style. *In this exceptional case I invite you to cite to/quote from this eprint in the event of discrepancies between the two versions. This is the author's favoured version.*

The Logic of Excuses and the Rationality of Emotions

JOHN GARDNER*

1. The challenge

Sometimes emotions excuse. Fear and anger, for example, sometimes excuse under the headings of (respectively) duress and provocation. Although most legal systems draw the line at this point, the list of potentially excusatory emotions outside the law seems to be longer. One can readily imagine cases in which, for example, grief or despair could be cited as part of a case for relaxing or even eliminating our negative verdicts on those who performed admittedly unjustified wrongs. To be sure, the availability of such excuses depends on what wrong one is trying to excuse. No excuse is available in respect of all wrongs. Some wrongs, indeed, are inexcusable. This throws up the interesting question of what makes a particular emotion apt to excuse a particular wrong. Why is fear, for example, more apt to excuse more serious wrongs than, say, pride or shame? This question leads naturally to another. Why are some emotions, such as lust, greed, and envy, apparently not apt to furnish any excuses at all? Can one not be overcome by them? Can they not drive one to wrongdoing as readily as fear and grief? Or is that not the point?

* Professor of Jurisprudence, University of Oxford. This paper was written while I was on leave at Princeton University in Spring 2008 and at the Australian National University in Summer 2008. Many thanks to both institutions for providing congenial working conditions, and to colleagues and students at both institutions for helpful criticisms and suggestions.

Is the point that these emotions, even if no less powerful than their potentially excusatory counterparts, are less defensible?

Here we already encounter a divide between two competing ways of thinking about emotions.[1] On one view, there is nothing to be said about the reasonableness or unreasonableness of our emotions, but only about their power to overwhelm us, and thereby partly or wholly to exempt us from the expectation of reasonableness that would aptly apply to us in their absence. On a rival view, emotions are themselves subject to appraisal as reasonable or unreasonable and hence can be included among the aspects of our lives in respect of which we are aptly expected to be reasonable. Corresponding to these two ways of thinking about emotions, there are two main ways of thinking about their excusatory force, which in turn invoke two rival accounts of excuses more generally. On one account, associated with J.L. Austin and H.L.A. Hart, those who offer excuses are thereby calling into question their responsibility for their actions, seeking partial or total exemption from the expectation of reasonableness that would otherwise apply to them as responsible agents.[2] According to the rival account, which I favour and which is largely favoured by the common law, those who offer excuses are, on the contrary, asserting their responsibility for their actions.[3] They claim not only to be subject to, but also to have lived up to, the normal expectation of reasonableness applicable to responsible agents. In the case of (what we might call)

[1] These views correspond approximately to those labelled 'mechanistic' and 'evaluative' by Dan Kahan and Martha Nussbaum in 'Two Conceptions of Emotion in Criminal Law', *Columbia Law Review* 96 (1996), 269, although in some ways Kahan and Nussbaum polarize the two views more than they need to do and more than I would be inclined to do.

[2] Austin, 'A Plea for Excuses', *Proceedings of the Aristotelian Society* 57 (1956), 1; Hart, 'Legal Responsibility and Excuses' in Sidney Hook (ed), *Determinism and Freedom in the Age of Modern Science* (New York 1958).

[3] See my papers 'The Gist of Excuses' and 'The Mark of Responsibility', both reprinted in my collection *Offences and Defences* (Oxford 2007).

emotional excuses, they say that their emotions were reasonable ones and that, since they acted on the strength of such reasonable emotions, their actions, although unjustified, are excused. They enjoy the duress excuse because and only to the extent that they acted in reasonable fear, the provocation excuse because and only to the extent that they acted in reasonable anger, and so on.

The second account of excuses – the one that portrays them as answering to reason - is subject to a challenge that helps to explain why so many people have been attracted to the first (Austin-Hart) account instead. If one claims to have lived up to the normal expectation of reasonableness applicable to responsible agents, then surely what one claims is no less than a justification? Where is the logical space for being reasonable – responding appropriately to reasons - without being justified? Unless we can find some logical space for this possibility, the second account does not so much explain excuses as explain them away. All so-called excuses turn out to be justifications.[4] The most tempting way to avoid this eliminative turn is to adopt the Austin-Hart account, which keeps excuses and justifications clearly segregated. By asserting one's reasonableness, according to this account, one always claims a justification. To claim an excuse is to go in the opposite direction, relying on the way in which one's emotions, or similarly powerful psychological forces, swamped one's rational faculties and thereby undermined the aptness of the expectation that one would react reasonably. True, this account preserves the distinction between justification and excuse only at the price of abandoning the other distinction we mentioned, the distinction between excuses and denials of responsible agency. But for many this price is the easier one to pay. Indeed for some it is not even a price. What turns, some

[4] A clear articulation of this argument, focusing on duress, is Peter Westen, 'An Attitudinal Theory of Excuse', *Law and Philosophy* 25 (2006), 289 at 348-50. I sketched a response to Westen in my 'Reply to Critics' in *Offences and Defences*, above note 3, 256-8. The present paper expands on this response.

wonder, on the distinction between excuses and denials of responsibility? Aren't both simply ways of securing an elimination, or at least a relaxation, of a negative verdict? Whereas something clearly turns on the distinction between justifications and excuses. Justified actions are to be welcomed, or at least endorsed - are they not? - while excused ones are still to be regretted and discouraged even though their agents are shielded from negative appraisal in the light of them.

I have various doubts about these last suggestions concerning what is as stake in the debate.[5] But for present purposes, I want to focus on the challenge that preceded them. Can we find logical space for excuses that answer to reason – that require reasonableness on the part of the agent – without collapsing excuses into justifications? I believe that we can. It is true that when one claims an excuse for what one did, one also claims to have been justified. But one does not claim to have been justified in what one did. On the contrary, one concedes that one was unjustified in what one did. One claims to have been justified in something else. In the case of emotional excuses, one claims to have been justified in one's emotions. The action is excused because the emotion in the thrall of which one acted is justified. Or – to put the same point in terms of reasonableness – it was an unreasonable action in the grip of a reasonable emotion. It was an excessive retaliation, for example, in reasonable fury at continual belittling. Or a premature strike in reasonable terror of impending attack. This is not to be confused with the idea that excuses are somehow partial or incomplete justifications. A wrong is part-justified only if there are reasons, albeit not sufficient reasons, for its commission. Yet a wrong may be fully excused, on the view I am advancing, even though there was no reason at all to commit it. What count under the excuse heading

[5] See 'In Defence of Defences' and 'Justifications and Reasons', both included in *Offences and Defences*, above note 3.

are not the reasons in favour of one's action but the reasons in favour of one's being in the condition in which one was driven to perform that action, the reasons in favour, for example, of one's being as despairing or as grief-stricken as one was.[6]

Unfortunately, this way of answering the original challenge invites two new challenges of its own.

First, what kind of connection is supposed to be represented by 'on the strength of', 'in the thrall of', 'in the grip of', and other locutions that I have so far used, apparently interchangeably, to express the connection between emotion and action? How, in cases of emotional excuse, is one's *ex hypothesi* justified emotion supposed to contribute to one's *ex hypothesi* unjustified action? Is the contribution the same as that which obtains when one acts 'on the strength of' a reason of which one is aware? If so, should we think of the relationship between emotion and action as akin to the relationship between belief and action, and should this in turn lead us to think of emotion as prone to mislead rationally in something like the way in which belief is prone to mislead rationally? Or could it be that by these various locutions I am at least sometimes helping myself to the rival idea that the contribution of emotion to action is non-rational, an alternative causal process that somehow ousts or overrides or jams the normal relationship between reasons of which one is aware and actions that one performs for those reasons? The latter idea chimes with the familiar Austin-Hart view of excuses. If I am forced to help myself to it in the course of articulating my supposedly contrasting view of excuses, then that suggests a dangerous instability in how I think about excuses, and not only in how I think about emotions.

[6] Cynthia Lee interestingly argues that, in the criminal law, provocation ought to be recognized as an excuse only when it is also a partial justification. Her discussion sheds light on the distinction. See Lee, *Murder and the Reasonable Man* (New York 2003), esp ch 10 (which also helps to illuminate and illustrate the idea of an unreasonable action out of a reasonable emotion).

Second, whatever 'on the strength of' (etc.) means, how can there possibly be reasons for emotions that are not equally reasons for the actions that one performs on the strength of them? If an emotion leads one to act unreasonably, how can it truly be a reasonable emotion? Isn't an emotion to be judged, if it is to be judged at all, by its contribution to action, and isn't a reasonable (and hence justified) emotion therefore simply one that motivates a reasonable (and hence justified) action? I will call the view that an emotion is to be judged by its contribution to action the 'practical conception' of emotion. The thesis that the practical conception may seem to support, by which a reasonable (and hence justified) emotion is simply one that motivates a reasonable (and hence justified) action, I will label the 'no difference thesis'. According to the no difference thesis any reason that there may be in favour of or against any emotion also favours or disfavours, to the same extent, the very action that the emotion motivates one to perform. Thus if there is a reason for me to be so afraid that in my fear I run away, that reason is necessarily also a reason, *pro tanto*, for me to run away. If this thesis holds, then my answer to the original challenge (where is the logical space for excuses as opposed to justifications?) obviously fails.

In this essay I will be concerned mainly with the second of these fresh challenges, which I will discuss in some detail. I will begin by refuting the no difference thesis while maintaining the practical conception of emotion (section 2). I will then proceed to reject the practical conception itself (section 3). By these two steps I will first open up, and then enlarge, the logical space available for emotional excuses on the model of excuses that I favour. As for the first of the new challenges – the idea that I am unofficially helping myself to a rival view of emotions, and thereby to a rival view of excuses – I will limit myself to some brief and rather peremptory remarks (section 4). Until then I will keep this challenge in the background by largely avoiding the offending expressions ('in the thrall of', 'in the grip of', etc.) in

favour of the bland expression 'motivated by' and its cognates: we are motivated by emotions; emotions motivate us.

2. Refuting the no difference thesis

By the no difference thesis, to repeat, a reasonable (and hence justified) emotion is simply one that motivates a reasonable (and hence justified) action. This thesis may seem to give emotions an excessively narrow, because purely instrumental, role in our practical lives. But this impression is misleading. The no difference thesis is consistent with the proposition that the intrinsic value of actions may be affected by the emotions that they manifest (=exhibit, express). It is worth noticing, however, that an action can only manifest an emotion that also motivates it. I cannot exhibit my frustration or excitement in my actions except by acting out of frustration or excitement. So the question is whether, when I do this, the intrinsic value of manifesting the emotion can contribute to its justification on top of its instrumental value as a motivation for the very action that manifests it. The answer can only be negative. Actions inherit intrinsic value from the emotions that they manifest only if the emotions in question are reasonable ones. It follows that the reasonableness of the emotion needs to be determined independently of the intrinsic value of manifesting it. Since the no difference thesis is a thesis about the reasonableness of emotions, its soundness is untouched by the intrinsic value of manifesting reasonable emotions in action.

This is not to say that there can be no value in manifesting one's unreasonable emotions in action. Sometimes, after all, by manifesting one's emotion in action one contains or eliminates it ('gets it off one's chest'). And the more unreasonable an emotion, *ceteris paribus*, the more reason one has to contain or eliminate it. By this line of thought, however, an unreasonable emotion still makes a purely instrumental contribution to the case, such as it is, for its own manifestation. For the *ex hypothesi* unreasonable

emotion has the following thing going for it, if nothing else. At least it motivates one to get it off one's chest. At least it tends to attenuate its own perpetuation. And this fact can readily be accommodated by the no difference thesis. The no difference thesis merely adds: Insofar as it is a redeeming feature of the unreasonable emotion that it tends to attenuate its own perpetuation, this fact weighs against the unreasonableness of the emotion to the same extent as it weighs against the unreasonableness of the action that manifests the emotion. In this respect, as in all others, the (un)reasonableness of an emotion, according to the no difference thesis, varies with the (un)reasonableness of any action motivated by it.

One does not, however, need to wrestle with such convoluted applications of the no difference thesis in order to uncover the most basic error underlying it. For it is an everyday occurrence that one has a reason to be motivated to perform an action that is not a reason to perform that action. This is because one's motivation to perform a particular action on a particular occasion typically has various consequences other than the performance of that very action on that very occasion. Most commonly and most importantly, one's motivation to perform a particular action on a particular occasion also motivates one to perform similar or related actions on other occasions. Often, one could not lose the motivation on one occasion without also losing it on other occasions. That being so, one's motivation must be judged in the light of its impact on one's actions across the range of those linked occasions, not only on the present occasion. I may be overmotivated to ϕ (escape, retaliate, resist, apologise, assist, explain …) for present purposes but that may be the price that has to be paid for my being correctly motivated to ϕ in the generality of cases in which, thanks to the roles that I normally occupy and the situations that I normally confront, my ϕing is exactly what is called for. Thus my motivation to ϕ may be the optimal one for me to possess even though the particular ϕing that it now motivates in me may be suboptimal. Putting the

same point in the idiom of reasons, the reasons I have to (continue to) be motivated to ϕ, such that I ϕ right now, may exceed (and might even dwarf) the reasons for me to ϕ right now. My motivation may therefore be reasonable (and hence justified) even though my action on the strength of it is not.

This point was most famously made by Robert Adams, who used it to draw a distinction between two rival ways of understanding and applying the Benthamite principle of utility. Adams illustrated the distinction with the example of Jack:

> Jack is a lover of art who is visiting the cathedral at Chartres for the first time. He is greatly excited by it, enjoying it enormously, and acquiring memories which will give him pleasure for years to come. He is so excited that he is spending much more time at Chartres than he had planned, looking at the cathedral from as many interior and exterior angles, and examining as many of its details, as he can. In fact, he is spending too much time there, from a utilitarian point of view. He had planned to spend only the morning, but he is spending the whole day; and this is going to cause him considerable inconvenience and unpleasantness. He will miss his dinner, do several hours of night driving, which he hates, and have trouble finding a place to sleep. On the whole, he will count the day well spent, but some of the time spent in the cathedral will not produce as much utility as would have been produced by departing that much earlier. At the moment, for example, Jack is studying the sixteenth to eighteenth century sculpture on the stone choir screen. He is enjoying this less than other parts of the cathedral, and will not remember it very well. It is not completely unrewarding, but he would have more happiness on balance if he passed by these carvings and saved the time for an earlier departure. Jack knows all this, although it is knowledge to which he is not paying much attention. He brushes it aside and goes on looking at the choir screen because he is more strongly interested in seeing, as nearly as possible, everything in the cathedral than in maximizing utility. This action of his is therefore wrong by act-utilitarian standards, and in some measure intentionally so. And this is not the only such case. In the course of the day he knowingly does, for the same reason, several other things that have the same sort of act-utilitarian wrongness.
>
> On the other hand, Jack would not have omitted these things unless he had been less interested in seeing everything in the cathedral

than in maximizing utility. And it is plausible to suppose that if his motivation had been different in that respect, he would have enjoyed the cathedral much less. It may very well be that his caring more about seeing the cathedral than about maximizing utility has augmented utility, through enhancing his enjoyment, by more than it has diminished utility through leading him to spend too much time at Chartres. In this case his motivation is right by motive-utilitarian standards, even though it causes him to do several things that are wrong by act-utilitarian standards.[7]

The example is not only of significance to utilitarians.[8] It is of significance to anyone who espouses the no difference thesis. For it shows, if it shows anything, that a motivation may be right (=reasonable, justified) even though some of the acts that it motivates are wrong (=unreasonable, unjustified).

One worry about the example is this.[9] Why can't Jack act for one reason in the morning (say, for the sake of sublimity) and another in the afternoon (say, for the sake of utility)? Why can't he vary his motivations to suit the demands of the situation, and

[7] Robert Adams, 'Motive Utilitarianism', *Journal of Philosophy* 73 (1976), 467 at 470-1.

[8] For a less narrowly utilitarian deployment of the same idea, see Derek Parfit, *Reasons and Persons* (Oxford 1985), 27-8. Parfit notes his agreement with Adams at 505.

[9] A different worry, urged by Fred Feldman against Adams and by Bart Gruzalski against Parfit, is that the principle of utility (being a moral principle) is concerned only with the comparative assessment of alternative courses of action that are available to the agent at the time of acting. *Ex hypothesi*, the alternative of leaving Chartres on time is not available to Jack when the time to leave arrives, given his motivations. So there is no act-utilitarian verdict on it. So act utilitarianism cannot reach a different verdict on it from motive utilitarianism. I will not consider this objection here as it muddles the conditions of moral responsibility in a way which sets it severely at odds with the presuppositions of my discussion, notably by leaving us with nothing to excuse. See Feldman, 'On the Consistency of Act- and Motive-Utilitarianism: a Reply to Robert Adams', *Philosophical Studies* 70 (1993), 201; Gruzalski, 'Parfit's Impact on Utilitarianism', *Ethics* 96 (1986), 760 at 771-777.

thereby meet the demands of act-utilitarianism (or more generally, reasonable action) throughout? The answer is that we aren't talking about Jack's reasons for acting when we talk about his motivations. We are talking about what Jack cares about, what moves him to attend to some reasons for acting more than to others in a way that need not be a function of (even what he knows to be) their comparative rational force alone. Jack, we are told, is an art-lover and an enthusiast for Chartres. Adams is assuming, I think plausibly, that such loves and enthusiasms[10] cannot but infect a range of Jack's actions across time. By their nature, loves and enthusiasms have some measure of continuity. Of course it does not follow that their course is predictable. They may wax and wane, or go into lulls. Yet it is inconsistent with having loves and enthusiasms, as opposed to merely feigning them, that one turns them on and off to suit the demands of the situation. That is why Jack's morning and Jack's afternoon should be thought of as linked occasions, in the sense that I already specified above. They are linked occasions in the sense that if Jack could switch off his love of art or his enthusiasm for Chartres as motivations in the afternoon when they distract him into various suboptimal actions, he would not have had the same love and enthusiasm in the morning, when they motivate him to various optimal actions that would otherwise be lost to him.

[10] In the sense that matters here, an enthusiasm, like a love, is always *for* something. It is possible to be an enthusiast for particular pursuits at the expense of others. So one can also have enthusiasms (plural). In a different (and here irrelevant) sense, enthusiasm (always singular) is a personality trait. A person can be generally enthusiastic in the way that she can be generally optimistic, sullen, alert, lively, moody, passive, etc.

Now love and enthusiasm are not emotions.[11] There are four interconnected features of emotions that they lack. First, love and enthusiasm, unlike emotions, lack any distinctive affective ingredient. Although it is possible to have emotions without feeling them – to be afraid without feeling afraid, to be ashamed without feeling ashamed, to be envious without feeling envious, and so on - each emotion has its distinctive affective ingredient - its own distinctive pleasure or pain[12] - experience of which is part of feeling the emotion. Love and enthusiasm, however, are characterized by affective diversity, even disarray. The point is most striking with interpersonal love. Insofar as it is felt, it is, as they say, a rollercoaster ride. There is the misery of being apart from the person one loves, the joy of her return, the delight one takes in her pleasures and triumphs, the jealousy of time she spends with others, the pride in her achievements, the fear of rejection, the grief of losing her, and so on. That love is so rich with emotional possibility, negative as well as positive, is one reason why it is so easily mistaken for an emotion. But it is not an emotion, for its diverse affective possibilities are inherited from

[11] For a robust defence of this view, as it applies to love, see O.H. Green, 'Is Love an Emotion?' in Roger Lamb (ed), *Love Analyzed* (Boulder 1997), 209. Others, such as George Pitcher, in 'Emotion', *Mind* 74 (1965), 326, and D.W. Hamlyn in 'The Phenomena of Love and Hate', *Philosophy* 53 (1978), 5, maintain that love is an emotion only by allowing that it is a deviant case. All three of these authors emphasize the third and fourth features that I mention below, although Green also relies on aspects of the second.

[12] Aristotle thinks about affectivity in terms of pleasure and pain: *Nicomachean Ethics* $1105^{b}21$ff. This can be misleading, and has indeed misled many modern writers on the emotions. We sometimes speak of pleasure and pain to identify raw sensations. Those with chronic back pain or a pleasurable sensitivity in their feet cannot be criticized for finding pain or taking pleasure in the wrong things. Their pleasure or pain – what we might call Benthamite pleasure or pain – has no object and does not answer to reasons. In this respect Aristotelian pleasures and pains (for this is how Aristotle thinks of pleasures and pains more generally: *Nicomachean Ethics* $1175^{a}18$ff) could not be more different. They have objects and answer to reasons.

other emotions that are, so to speak, its symptoms. This is not to say that there are no affective experiences that are associated uniquely with love. Maybe there are uniquely love-symptomatic emotions, such as the affection that sometimes wells up in one, bringing a tear to one's eye, when one catches sight of a beloved person who is peacefully asleep. However one underestimates love if one confuses it with this kind of affection, or with any other singular emotion. Different people's loves, as well as different loves of the same person, and especially different types of love (parental love, romantic love, puppy love, love of humanity, etc.) occupy different affective ranges, some more negative, some more positive, some more erratic, some more constant, some more intense, some more profound, some more contained, some more all-pervasive, and so on. The same, although usually less vividly, is true of enthusiasms, which may likewise be beset by disappointment as well as delight, envy as well as fellow-feeling, shame as well as pride, and so forth.

Second, as just foreshadowed, love and enthusiasm lack a defining orientation. Every emotion has either a positive or a negative orientation, which colours its distinctive affective ingredient (as positive or negative) as well as helping to explain its characteristic constituent desires for action. One can, of course, have mixed feelings, simultaneously beset by both positive and negative emotions towards the same object, taking both pleasure and pain in it. Yet there is no such thing as a (singular) mixed emotion. Why? The most elementary ingredients of each emotion, present even in the limit cases of that emotion, are a belief and a wish. One believes that P and one *either* wishes that P (positive orientation) *or* wishes that not-P (negative orientation) – but not both.[13] So, for example, Jack is irritated about his departing late only if (a) he believes that he

[13] For more investigation of this bivalent (but never ambivalent) aspect of emotions, see Robert Gordon, *The Structure of Emotions* (Cambridge 1987).

departed late and (b) he wishes that this were not so. And Jack fears driving through the night only if (a) he believes that he might have to drive through the night and (b) he wishes that this were not so. Whereas Jack is thrilled to have spent the day at Chartres only if (a) he believes that he spent the day at Chartres and (b) he still wishes that it have been so. In each of these examples the belief supplies the object of the emotion and the wish supplies what I am calling its orientation. Love and enthusiasm do not fit the model. You may find this suggestion bizarre. Surely love and enthusiasm have objects and positive orientations towards them if anything does? But you are thinking here of a positive orientation towards (=an attraction or allegiance to) a beloved person or thing or towards an enthusing pursuit, not a positive orientation towards (=a wish about) a given proposition P.[14] Once one starts to catalogue the common emotional symptoms of love, of course, one also starts to find some relevant Ps with wishes attached. One is miserable at the absence of one's beloved (one (a) believes that she is absent and (b) wishes that it were not so); one is glad of her existence (one (a) believes that she exists and (b) wishes that it be so); one is afraid of her disapproval (one (a) believes that she might disapprove of one and (b) wishes that it were not so); and so forth. But none of these various Ps is itself the object of one's love. Love does not take a proposition as an object, and none of these orientations towards propositions, therefore, is the orientation of love. Again, the same can be said of enthusiasm. In the relevant sense, love and enthusiasm do not have any defining orientations. What have defining orientations are the various emotions that are often but not always symptomatic of them.

Thirdly, unlike love and enthusiasm, each emotion has special answerability to its own distinctive built-in rationale. As

[14] This helps to explain how it is that, while vicarious love is impossible, there can be vicarious emotions (e.g. fear or embarrassment or anger or joy or pride on someone else's behalf, including contagiously).

well as a belief *that* P, a normal emotion is partly constituted by a belief *about* P, viz. a belief that there is in P something that constitutes a reason for one's wishing that P or that not-P.[15] This secondary belief (as I will refer to it later) plays the major role, in combination with affective differences, in distinguishing among different emotions which have the same object and orientation. Different emotions, in other words, respond to different (supposedly) reason-giving properties of their objects.[16] For example, in the normal (non-phobic) case of fear one believes that there is in P something dangerous or threatening, which is one's fear-defining reason to wish it were not so. So Jack (non-phobically) fears driving through the night only if (a) he believes that he might have to drive through the night, and (b) he wishes that this were not so, and (c) – here is the extra condition that makes his fear non-phobic - he believes that there is something dangerous or threatening about his driving through the night, which is his reason for his wishing it were not the case that he might have to do it. There is no similar ingredient in love or enthusiasm. This is not to say that love and enthusiasm do not answer to reasons. It is merely to say that there is no distinctive reason that either of them answers to, because there is no distinctive reason, belief in which is an ingredient of them, nothing paralleling the supposed danger or threat in the case of fear. Once again, you may baulk at this. How about a belief in the lovability of what is loved or a belief in the enthusingness of what evokes enthusiasm? These may indeed be ingredients of normal love and enthusiasm, but they do not parallel the secondary belief that is an ingredient of a normal emotion, such

[15] For more detailed discrimination of the two beliefs, see Gabrielle Taylor, 'Love', *Proceedings of the Aristotelian Society* 76 (1976), 147 at 147-8.

[16] Although sometimes (e.g. as between pity and *Schadenfreude*) it is not clear whether two emotions differ in what they take to be the reason-giving property or only in their orientation, i.e. the way in which the same property is responded to. This is a problem with the individuation of properties.

as the belief in danger or threat that is an ingredient of fear in non-phobic cases. Rather, they parallel the belief in *frighteningness* that characterizes the object of fear in the eyes of the fearer even in phobic cases. Frighteningness we could call a placeholder or formal property; it is *whatever* property something is believed to have which, for the fearer-believer, is its fear-evoking property.[17] The property could be, e.g., the property of being a spider or the property of being a small enclosed space. It does not yet disclose a reason, an intelligible reason, for being afraid. This placeholder property of frighteningness resolves, in non-phobic cases, into a property of supposed danger or threat, which does disclose an intelligible reason for being afraid, and shows the fear to be a non-phobic one. There is no like resolution of the placeholder properties of lovability and enthusingness that would serve to distinguish pathological from apt cases of love and enthusiasm. For unlike an emotion neither love nor enthusiasm is part-constituted by a belief, on the part of the person experiencing it, in any distinctive reason for its own existence.

Finally, although love and enthusiasm are like emotions in answering to reasons, they are not so completely answerable to reasons as emotions are. Love and enthusiasm have important roles in human life partly because, when we put aside our loves and enthusiasms, there are so many rationally appealing things for each of us to do, and so few opportunities to combine them in a single life. Incommensurably appealing alternatives constantly rival each other for our attention. In the face of this rivalry we reasonably (and inevitably) adopt attitudes towards particular people or things or pursuits that do not defy, but are equally not dictated by, the independent rational appeal of those people or things or pursuits. Because there is a reason to have some such attitudes – without them there is no way to chart a course

[17] See Ronald De Sousa, *The Rationality of Emotions* (Cambridge, Mass. 1987), 122.

through life – once we have these attitudes, the fact that we have them qualifies as an extra reason for acting on top of those that justified our having them. 'I love A' therefore cites a valid reason for my bestowing special attention on A.[18] It is no objection to my bestowing special attention on A because I love A that I also need some independent reasons to love A. Most people, I suspect, have sufficient independent reasons to love whomever they love. But these reasons would also justify their loving any one of numerous similar people whom, quite by historical accident, they do not love. Only once one actually loves does one have the extra reason one needs – the fact that one loves - to justify one's bestowing special attention on just one (or a very small number) of these people to the exclusion of the others.[19] Emotions do not perform the same selecting or sorting role. One can therefore be expected to have a more complete rational explanation for one's fear, irritation, pity, frustration, or embarrassment (or one's lack thereof) than one can be expected to have for one's love. One can reasonably be called upon to explain what makes *that* so frightening, and *that* so irritating, and

[18] Bernard Williams thought that attempts to present love as fully answerable to reason (e.g. under the principle of utility) were alienating, because they made the complete reason (e.g. for looking after A) not 'I love A' but 'I love A and the principle that one first looks after those whom one loves is an efficient way of dividing up labour' (or a demand of duty or something like that). This is Williams' famous 'one thought too many' objection: 'Persons, Character, and Morality' in Amelie Rorty (ed), *The Identities of Persons* (Berkeley 1976), 197 at 214. It would be better thought of, however, as his 'one reason too few' objection. It is really a defence of the view that loving A can be an additional reason in its own right for acting in certain ways towards A, on top of the many reasons, such as the one mentioned above, that attach auxiliary salience to one's loving A. Only when that is cleared up is there the further question of which of these various reasons one is to act on and when.

[19] For exploration of the same point in relation to the attitude of respect, understood in a way that is somewhat akin to trust, see Joseph Raz, 'Respect for Law' in his *The Authority of Law* (Oxford 1979).

that so pitiable, and *that* so embarrassing, and so on, in the face of the many other things to which one might instead be paying attention, or by which one might instead be affected. And by the same token one cannot cite one's resentment or fear or shame or grief – unlike one's love or enthusiasm - as a reason in its own right for what one does. An emotional explanation for one's action is not yet a divulging of one's reason for so acting, but rather an explanation that points to the existence of (yet-to-be-divulged) reasons. Saying 'I did it because I was angry' (unlike 'I did it because I love her') properly attracts the rejoinder 'Yes, but what were your *reasons*?' meaning something like 'What were the reasons for your anger that would also help to explain your having acted as you did in anger?'

None of this detracts from the application of Adams' argument to emotions. Indeed the last point reveals that it applies in a more straightforward way to emotions than it does to love. For Adams ignores the fact that Jack's love of art is not fully answerable to reasons. Instead he assumes, *arguendo* and contrary to fact, that Jack's love of art is fully answerable to reasons, more specifically that it is fully answerable to the Benthamite principle of utility. That is because his point is about a second role for loves and enthusiasms. His point, put simply, is that loves and enthusiasms, even if *arguendo* they are not needed to fill in gaps in rationality, can still be rationally efficient. They can distort our motivations for good as well as for ill. There is always the question of whether one's loves and enthusiasms do more good than ill across a range of linked occasions. And exactly the same question can be asked more simply of emotions. Given that they sometimes incline one towards unreasonable actions and away from reasonable actions, is that the price that must be paid for their also, on linked occasions, inclining one towards reasonable actions and away from unreasonable ones? Anger, fear, guilt, *Schadenfreude*, envy, joy, pride, and jealousy, just like love and enthusiasm, generally affect a range of actions over time, and in assessing a particular emotion as it contributes motivationally to a

particular action, we generally need to consider motivational contributions that the same emotion makes to other actions too. Although suboptimal on one occasion, an emotion may be optimal when assessed across a range of linked occasions.

An example: Fired with indignation at some patronising remark by her boss, and determined to put him right, Jill may take a more forthright and effective line in this morning's negotiations, be more careful in criticising her own staff at this afternoon's team meeting, and finally deal with the backlog of email in her inbox before going home. 'I'll show him,' she may well be thinking. All of this may mark an improvement, albeit perhaps short-lived, in Jill's work habits. Yet her indignation may also have led her to repeat a mean-spirited and unfunny joke about her boss at lunchtime, a low-point in her behaviour for the day. She would have been less inclined to tell the joke only if she had been less inclined to show her boss her mettle, for it was the same indignation that spurred her to both. The joke was a price that had to be paid for the day's many achievements. Jill's indignation was then reasonable (justified) even though her telling that joke in indignation was unreasonable (unjustified).

The same worry we had about Adams' example of Jack may reassert itself more powerfully in the case of Jill. For surely she can vary her motivations to suit the demands of the occasion? The explanation of why her indignation is optimal (to the extent that it is) seems to be precisely that its expression in action is controlled, measured, contained, disciplined. The characteristic desire of the indignant person, you may say, is to ensure that the person towards whom she is indignant should not get away with it, i.e. that he is punished for, or at least does not profit from, his wrongs. In Jill's story the indignation issues only once in an action of this type, and this is the very action that is paraded as unjustified, the 'low point' of Jill's day. Isn't the main lesson of this that indignant people, and others in the throes of emotions, are properly expected to exercise self-control in respect of the expression of their emotions in their actions? This chimes, you

may say, with the thought that emotions answer fully to reason. Jill believes that her boss patronised her and this is the reason that she takes herself to have for her desire to punish him. Can she not be expected to tailor her indignant actions to the force of this reason, at any rate to the force of this reason as she believes it to be? Can she not be expected to note the inflationary motivating effect of her accompanying wishes, desires, and affects, and to control them as the situation demands?

Within limits, no doubt. But the objection, as I just formulated it, does not seem to respect the limits.[20] It misses the important point that the relevant belief – the belief in a reason to act - is itself distorted by the fact that it is an indignant belief. The fact that it brings with it the characteristic wishes and desires and affects of indignation gives a salience or prominence to one's P-given reason, in one's indignant mind, that reduces one's scope

[20] It also, incidentally, oversimplifies the characteristic desire associated with indignation, which is a desire to show the object of one's indignation the error of his ways (or to 'put him right'). This can include (literally) demonstrating to him that he was mistaken or (metaphorically) 'teaching him a lesson' by punishing him. These two characteristic ways of expressing one's indignation are not always compatible. Jill can best demonstrate to her boss that he is wrong about her by doing her work well, but quite possibly the best way to punish him is by deliberately doing the same work badly. One implication of this is that one cannot rely on the fact that the indignation is justified to help one to justify the punishment, unless one is also willing to explain why it is alright to express one's indignation in a punitive way rather than in some other (not deliberately painful or deprivatory) way. The best one can hope for without that explanation is an excuse, not a justification, for one's punitive actions. It follows that the problem of the justification of punishment is postponed, not solved, by reliance on the aptness of indignation (and similar moralistic emotions). This is a problem in, for example, Samuel Pillsbury, 'Emotional Justice: Moralizing the Passions of Criminal Punishment', *Cornell Law Review* 74 (1988), 655, and some parts of the debate between Jeffrie Murphy and Jean Hampton in their *Forgiveness and Mercy* (Cambridge 1988). I highlight some implications of this point in 'Crime: in Proportion and in Perspective' in my *Offences and Defences*, above note 3.

for enkratic, or self-controlled, action.[21] It reduces that scope by reducing the distance between what one takes to be the force of the reason that one takes oneself to have and the force of one's desire to conform to it. This was equally true, although we glossed over it, in the case of Jack. Jack, recall, 'attend[s] to some reasons for acting more than to others in a way that is not a function of their comparative rational force alone.' And he is not so much driven by his love of art, remember, as 'distracted' by it. His love of art infects not only his desire to engage with Chartres but also his beliefs about his reasons to engage with Chartres, such that there is no bare appreciation of those reasons, untainted by his love of art, to which he can be expected to tailor his desires by any exercise of self-control, however impressive. The same is true with emotions, sometimes dramatically. For in the grip of an intense emotion we are contending additionally with the combined force of various affects (and sometimes even physical sensations) that conspire to focus our attention dramatically upon some reasons at the expense of others, that tend to make us believe in the reasonableness of what we intensely want to do, or (in other words) that tend to make our desires, in our eyes, more or less thoroughly self-vindicating. Emotional action, in other words, is not typically akratic action. One does not typically judge ϕing to be wrong but persist in ϕing under the influence of one's emotions. Typically, one judges ϕing to be right under the influence of one's emotions when, absent the emotions, one might judge ϕing to be wrong.[22] So there is limited scope for self-control in respect of one's emotion-motivated actions unless one first controls one's

[21] I ignore here the interesting question, raised by Richard Holton, of whether self-control is exactly the same as enkrasia. Holton, 'How is Strength of Will Possible' in Sarah Stroud and Christine Tappolet (eds), *Weakness of Will and Practical Irrationality* (Oxford 2003), 39 especially at 55-6 (n21).

[22] Emotions 'change people so as to alter their judgments': Aristotle, *Rhetoric* 1378^{a}20.

emotions. Which brings us back to the original question we are interested in: When it is unreasonable to have an emotion (such that it would be reasonable, if one could, to control it)?

Some people seem to think that the inevitable feedback effect of emotion upon one's judgments of right and wrong, and more generally upon one's beliefs about one's reasons for acting, yields a general practical case against emotions. Emotional judgment is clouded. If they are to act well, people should not be emotional, or should be less emotional than they are. The Adams argument, as developed here, squarely undermines this view. What the Adams argument shows is that there is no general or default answer to the question of whether a (so to speak) sober appreciation of reasons for action is more reasonable than an emotionally-charged appreciation of those same reasons for action. Being cool, calm, and collected is just another place on the emotional map, with no special claim to rational efficiency.

If consideration of a single emotional episode such as Jill's does not bring this point home for you, then think instead of emotional dispositions. It is not possible for a sane human being not to have a range of emotional dispositions, adding up to what is usually known as a temperament. Even one who is disposed to be unemotional has a temperament. We call him a cold fish. So one may reasonably want to know: What temperament should I have? Should I be a cold fish or a hothead? Stoical or stricken? A timorous beastie or a tiger? Sensitive or stolid? The answer varies depending on whether one is (or is thinking of becoming) an actor or an airline pilot, a therapist or a traffic warden, a parent or a football coach, an entrepreneur or an academic, a negotiator or a nurse, and so on. Although it is possible for one person with one temperament to move with success between these and other very different roles, different roles are often suited, optimally, to people with different temperaments. Even an actor may on one startling occasion be faced with the situation in which he is the last person still conscious on an aircraft in mid-air, and has to bring it in to land under radio instruction from the control

tower. In this delicate situation he may make mistakes that are owed, not to his want of skill as a pilot, but to his want of steeliness or *sang-froid*. If only he had been more cut out to be a pilot! Or perhaps not. Perhaps then we would never have had that intensely melancholic portrayal of Vanya or that uniquely bruising rendition of Lear. This story presents us with a large range of linked occasions, and for most of them our hero is cut from exactly the right cloth. But not, alas, for all of them. Like the rest of us – in defiance of the no difference thesis - he does occasional unreasonable (unjustified) things thanks to a pattern of emotional response that is reasonable (justified) given the roles that he normally occupies and the situations that he normally confronts. Unfortunately, this pattern of emotional response turns out to be less than optimal for emergency landings.

3. Beyond the practical conception

As promised, I have rejected the no difference thesis without yet challenging the practical conception of emotion that inspired it. According to the practical conception of emotion, recall, an emotion is to be judged entirely by its contribution to action. Even in Adams-style cases in which there are reasons for an emotion that are not reasons for a given action motivated by that emotion, the reasons under discussion are still practical reasons in the following sense. They are reasons for the emotion that derive from the value of *other* actions to which the emotion also contributes. This is how the practical conception of emotion survives the failure of the no difference thesis.

What else could these reasons for emotions be but practical reasons, you may ask? Practical reasons are usually contrasted with epistemic reasons, also known as theoretical reasons or reasons for belief. Emotions, as we saw, have beliefs among their ingredients. No doubt these beliefs answer to epistemic reasons. That much is unproblematic. On any credible view there is a subsidiary epistemic component in practical reasoning. One

always needs to judge how things are as part of judging how to react. This applies no less to emotional reactions than to any other reactions. The question is how the *rest* of an emotion – its various ingredients beyond its belief ingredients – could possibly answer to epistemic reasons. How could there be an epistemic reason, for example, for the element of *wishing* (that P were so, or that P were not so) that goes into every emotion? The idea seems inconsistent with the most familiar way of drawing the distinction between a belief on the one hand and an attitude such as a wish on the other. According to the most familiar way of drawing this distinction, beliefs have a 'mind-to-world' direction of fit: 'beliefs should be changed to fit with the world, not *vice versa*.'[23] Wishes, along with desires, have a 'world-to-mind' direction of fit: 'the world, crudely, should be changed to fit with our desires [and wishes], not *vice versa*.'[24] On this direction-of-fit analysis there could not possibly be epistemic reasons for wishes or desires, for if they answered to the 'should' of belief – the 'should' of epistemic reasons[25] - they would no longer qualify as wishes or desires. By the same token there could not possibly be epistemic reasons for emotions, where this is taken to mean that the reasons in question would favour the emotion's constituent wishes and desires as well as its constituent beliefs.

If it is also supposed to resolve into an explanation of the distinction between practical and epistemic reasons – an explanation of the difference between the two 'shoulds' that figure in the quoted formulations - the direction-of-fit analysis is flawed and needs to be revised. But before we revise it there is some good sense to be extracted from it. It is true that there is a

[23] Mark Platts, *Ways of Meaning* (London 1979), 257.
[24] Ibid.
[25] Here I am presupposing that 'should' needs to be analyzed in terms of reasons, not *vice versa*. This view has been challenged, I think unsuccessfully, by John Broome in a series of important papers starting with 'Normative Requirements' *Ratio* 12 (1999), 398.

class of reasons, exemplified by evidential and testimonial reasons for belief, which have the following character. They support or favour the tracking, in the way we think, of the way things are. If the way we think does not track the way things are, then so far as conformity with these reasons is concerned the problem is in the way we think, not in the way things are. So there is no case, based on these reasons, for changing the way things are so that they match the way we think; the only case is for doing the opposite.[26] I will call these reasons 'tracker reasons'. Not all reasons for belief are tracker reasons. The Spanish Inquisition infamously provided reasons to believe in God which were not tracker reasons. They were reasons to believe in God's existence irrespective of whether God existed. There is something anomalous about these non-tracker reasons for belief, to be sure. Yet they are reasons for belief all the same.

More salient for present purposes, however, is the converse possibility. Are there tracker reasons that are not reasons for belief? When we speak of the tracking of the way things are in the way we *think*, do we mean only in what we believe? No. The class of tracker reasons extends widely beyond reasons for belief. It includes certain reasons for moods (*ceteris paribus* one should be gloomy when things are going badly, pensive when there are things to think about, and so on) and certain reasons for attitudes such as love, respect and admiration (*ceteris paribus* one should love and respect and admire people in proportion to their good personal qualities). Most saliently, for present purposes, the class of tracker reasons includes certain reasons for emotions. They are reasons to adjust one's emotions to suit the way things

[26] For the modern origins of this explanation of tracker reasons, see Elizabeth Anscombe, *Intention* (Oxford 1957), 56. This passage by Anscombe is the one from which the direction-of-fit analysis of beliefs v desires, and thence of epistemic v practical reasons, has mainly been developed. See I.L. Humberstone, 'Direction of Fit', *Mind* 101 (1992), 59. However Anscombe's remarks do not support most of the familiar developments of them.

are – to get angry at an insult, to be afraid of a danger, to be embarrassed at an *exposé*, to grieve over a loss, to be proud of an achievement – but not to adjust the way things are apart from one's emotions. In particular, they are not reasons to engineer a suitable insult to match one's anger, to seek out a suitable danger to match one's fear, or such like. What matters is the tracking of the world by the emotion, not *vice versa*.

That is the measure of good sense in the direction-of-fit analysis, and it allows certain reasons for emotions, attitudes, and moods, to be lodged, alongside certain reasons for belief, on the non-practical side of rational life. The problem with the direction-of-fit analysis concerns what lies on the practical side of rational life. Practical reasons are not reasons to adapt the way things are to the way one thinks, or otherwise to bring world into conformity with mind. Practical reasons do not relate world to mind at all. They are reasons to adapt (or preserve) the way things are (including aspects of ourselves) in the service of value.[27] I have included the words 'including aspects of ourselves' as a reminder that our thoughts too may be touched (Spanish-Inquisition-style) by reasons of this kind. For the purpose of practical reasoning, our thoughts are not what the world is to be adapted to. Rather our thoughts, now including our emotions and beliefs and desires and so forth, are further things in the world that may need to be adapted in the service of value. As our discussion of the Adams argument in the previous section showed, practical reasons can be reasons for thoughts themselves, inasmuch as those thoughts contribute to actions. They need not be reasons for the further actions of bringing those thoughts into or out of existence, such as the fact that they are painful or pleasant thoughts. To forestall any misconceptions on this score,

[27] In this formulation, and more generally in thinking about the two types of reasons contrasted here, I was greatly assisted by Joseph Raz's 'Reasons: Practical and Adaptive' in David Sobel and Stephen Wall (eds), *Reasons for Action* (Cambridge, forthcoming).

I will abandon the term 'practical' to describe the reasons I am currently discussing. I will call them 'service reasons' instead. To summarise: service reasons differ from tracker reasons in the following way. Tracker reasons are reasons to adapt (or preserve) the way one thinks (including the way one feels, the attitudes one takes, etc.) so as to track the ways things are; service reasons, by contrast, are reasons to adapt (or preserve) the way things are (including the way one thinks, which in turn includes the way one feels or inclines etc.) in the service of value.

Normal reasons for belief (setting aside those strange Spanish-Inquisition-style reasons) are tracker reasons. They do not exist in the service of value. This is obscured by the inflationary way in which we sometimes talk about truth. As belief-capable beings, we are sometimes said to be truth-seekers. Truth, we sometimes say, is the built-in aim of believing.[28] So we are, and so it is. But these are simply grand ways of saying that all beliefs answer, in their nature, to the way things are, inasmuch as the way things are is capable of figuring in the content of beliefs. It does not mean that beliefs have a built-in aim in the way that, say, winning is a built-in aim of chess or being together is a built-in aim of marriage. For winning and being together are valuable apects of chess and marriage (or at least are taken or held out to be valuable by chess-players and people getting married). Whereas truth, in the sense in which the believer necessarily aims at it, is not a value (and need not be taken or held out to be of value by believers). There is a reason, when a tennis player is losing, for her to up her game so that in the end she wins. But there is no truth-directed reason, when people have false beliefs, to change the world just in order to make their beliefs true. That people in Britain today obstinately believe that there is more violent crime around them than there really is is no reason, not

[28] Or 'beliefs aim at truth', as Bernard Williams influentially states the idea in 'Deciding to Believe', in his *Problems of the Self* (Cambridge 1973), 136.

even a heavily defeated reason, to increase the amount of violent crime in Britain so as to increase the amount of true belief. On this ground, if on no other, we should not inflate truth with a capital 'T' and put it alongside Beauty and Justice on our list of noble ideals.[29] Truth, to repeat, is no more than the way things are, inasmuch as this is capable of figuring in the content of beliefs.[30] Truth lends no value to things being as they are or to anyone's believing them to be so.

We are all familiar with the existence of inconvenient or painful or otherwise regrettable truths. These explain our human *penchant* for deception, including self-deception. In analysing regrettable truths, we may be tempted to think that the intrinsic value of the belief *qua* true competes with its instrumental disvalue *qua* inconvenient, painful, etc. Then all reasons for and

[29] Consider John Rawls' famous analogy between truth and justice in *A Theory of Justice* (Cambridge, Mass 1971), at 3: 'Justice is the first virtue of social institutions, as truth is of systems of thought.' Although Rawls does not go as far as to capitalize the initial letter of 'truth', he gives succour to twin errors of inflation. First, since justice is a value, that truth is a value too (this inflates truth), and second, that in politics we need what truth lends to belief, viz. an overarching built-in objective (this inflates justice). As this suggests, we should be as cautious about justice with a capital J and beauty with a capital B as we are about truth with a captial T, but on different grounds.

[30] As opposed to the way things are *qua* mood-apt, emotion-apt, etc. To spell out the implication: Not all tracker reasons are truth-directed reasons. One may conjure up an 'emotional truth' to fill the space in the case of emotions that truth occupies in the case of beliefs: see Ronald de Sousa, 'Emotional Truth', *Proceedings of the Aristotelian Society Supplementary Volume* 76 (2002), 247; Mikko Salmela, 'True Emotions', *Philosophical Quarterly* 56 (2006), 382. But this proposal, taken as literally as these two authors want us to take it, is doubly misleading. First, it encourages the inflation of ordinary truth to capital-T status. Secondly, it overstates the closeness of the analogy between tracker reasons for beliefs and tracker reasons for emotions in ways that I will try to forestall below. These criticisms do not apply to those – such as Martha Nussbaum, *Upheavals of Thought* (Cambridge 2001), 46 - who say that emotions can be true since they *are* beliefs. This view falls at an earlier hurdle.

against believing, even truth-directed reasons, are presented as service reasons. But this analysis quickly breaks down. If there were no regrettable truths there would be nothing in the world to change, so there would be no service reasons at all.[31] It follows that the existence of service reasons depends on the existence of another class of reasons that are not service reasons. Truth-directed reasons for belief belong to (but do not exhaust) this class. They make possible a situation in which there is no value in believing something that one is nevertheless justified in believing, maybe even rationally compelled to believe. The belief brings no good to the world. It is nothing but a burden, a distraction, or a disappointment. There is not a single (service) reason in its favour. And yet there are overwhelming (tracker) reasons in its favour. There is, for example, decisive evidence and testimony that establishes its truth beyond reasonable doubt.

If we are familiar with this disturbing configuration in the case of beliefs, surely we are familiar with it all the more in the case of emotions. How often is one faced, for example, with an anger that is entirely fitting – it measures up perfectly to the indignity one suffered – and yet is entirely unproductive, a complete waste of time and energy, a hiding to nothing? Or how about a compassion that entirely befits the tragic events unfolding before one's eyes, but that brings in its wake only a sense of futility and a retreat into inertia? Or how about regrets of the Edith Piaf type: appropriate to '*le mal … qu'on m'a fait*' but demanding to be swept away in the cause of a better future ('*ça*

[31] Here I am not presupposing that all service reasons are reasons in favour of change. Some are reasons against it, i.e. in favour of the *status quo*. But unless service reasons for change are conceivable there can be no service reasons against it either. Where there are no conceivable alternative paths, reasons have no conceivable work to do, either for or against, and hence do not arise. To put it another way, every reason for something is necessarily a reason against something else, if only the absence of the first something.

commence avec toi')?[32] Again we may be tempted to think, in these cases, that the intrinsic value of the emotion *qua* fitting competes with its instrumental disvalue *qua* wasteful, destructive, etc. I will retrieve a grain of truth from this suggestion in a moment. But as it stands it is mistaken. The fittingness in question is not a value of the emotion, and the reasons for the emotion, *qua* fitting, are not service reasons. They are tracker reasons. They speak in favour of the emotion and lend it *ceteris paribus* rational propriety that is not based on its value, whether instrumental or intrinsic.

These tracker reasons, you will notice, are the very same reasons that I mentioned in my characterisation, in the previous section, of the secondary beliefs that belong to the constituents of (normal) emotions. We are talking about the (believed) danger or threat in the case of fear, the (believed) indignity or obstruction in the case of anger, the (believed) loss of someone or something beloved in the case of grief, the (believed) misfortune of others in the case of pity, the (believed) favour or gift received in the case of gratitude, and so on. Regarding oneself as having these reasons, as I said, is an integral part of experiencing the corresponding emotions in all but exceptional cases. You may wonder: How is it that we already encountered these tracker reasons back in the previous section, when at that point we were supposed to be working within the limits of the practical conception, i.e. granting the answerability of emotions to service reasons alone? The answer is that some tracker reasons for emotions can be represented, without too much strain, as roundabout service reasons: fear tracks danger or threat, for example, just because it is better that people in danger or under threat are motivated to flee. At any rate (adds the Adams argument) it is better that people in danger or under threat are motivated to flee in general, across a range of linked occasions,

[32] Michel Vaucaire and Charles Dumont, 'Non, Je Ne Regrette Rien' (1956), recorded by Piaf in 1961 and associated most closely with her.

even though it is not better that they are motivated to flee on this particular occasion (say, from a grizzly bear). One may accordingly come across the tracker reasons in discussions of emotions without noticing that one is doing so. They are all too readily assimilated to the discourse of service reasons. This is intriguing. Truth-directed tracker reasons would not so readily be assimilated to the discourse of service reasons in a parallel discussion of the rationality of belief. That is because service reasons for belief are so conspicuously odd. Epistemic pragmatism, although it has its adherents, is a stressful doctrine to maintain. One must constantly confront the counterintuitiveness of one's position, presenting the Spanish Inquisition, for instance, as a paradigmatic rather than an anomalous giver of reasons for belief in God. Yet our discussion of the rationality of emotions in the previous section did not require us to engage in complex contortions to maintain the practical conception against obvious doubts. There was no sense of having to explain anything away. It was quite natural to (mis)interpret the tracker reasons for emotion as yet more service reasons for emotion.

Why? Maybe the explanation lies in the following apparent disanalogy between the case of belief and the case of emotion. While there are service reasons as well as tracker reasons for and against belief, only the tracker reasons – the truth-directed reasons - are such that one can come to believe or not to believe, for those reasons, the very thing that they are reasons to believe or not to believe.[33] The most one can arrive at, for those reasons, is acceptance or non-acceptance of the thing that they are reasons to believe or not-believe, which falls well short of belief or non-belief in it. This is what gives us the sense that service reasons for and against belief are anomalous: they do not meet one of the normal conditions for something to qualify as a reason, viz. that

[33] See the classic discussion in Bernard Williams, 'Deciding to Believe', above note 28, at 147-51.

one can conceivably conform to it by following it.[34] Is the same true of emotions? Only asymmetrically. One cannot be angry or embarrassed or ashamed or afraid because it would be good to be angry or embarrassed or ashamed or afraid. One cannot be afraid, say, for the reason that it would be better to flee (a service reason), but only for the reason that something is dangerous or threatening (a tracker reason). On the other hand, within limits one can weaken and even eliminate one's emotions, unlike one's beliefs, for service reasons. Utterances such as 'getting irritated won't solve anything', 'no need to panic', and 'no use worrying about it' and '*Je n'ai plus besoin [des regrets]*' invoke service reasons against emotions that are supposed to be followed by their addressees in reducing or eliminating those very emotions. Such utterances need not be futile. Nor need they be interpreted as a kind of surreptitious therapy that purports to invoke service reasons but in fact merely uses that invocation as a device to distract their addressees from the relevant tracker reasons. Nor need they be interpreted as merely reasons to suppress emotions in the sense of keeping them to ourselves, not expressing them in action. In their role as reasons against the emotions, rather, they coexist on equal terms with invocations of tracker reasons, such as 'it's not that frightening', 'there's nothing to be ashamed of', 'where's the insult in that?' and '*C'est payé, … [le] passé*'.[35] Both kinds of reasons are available to guide us in the containment of our emotions even if not in the acquisition of them. Both can be followed, if unreliably and usually with slow results. Perhaps for that reason a discussion of emotions in which service reasons are

[34] I used to hold the stricter view that something not meeting this condition does not qualify as a reason at all. See John Gardner and Timothy Macklem, 'Reasons' in Jules Coleman and Scott Shapiro (eds), *The Oxford Handbook of Jurisprudence and Philosophy Law* (Oxford 2000). Further discussion with Timothy Macklem has persuaded me that this view was too strict.

[35] Again Williams' discussion is illuminating: see his 'Morality and the Emotions' in *Problems of the Self*, above note 33, 223–4.

treated as standard, like our discussion in section 2, does not seem so tortured as would a discussion of beliefs with the same feature. That, anyway, strikes me as a hypothesis worth testing.

We should not, however, allow it to mislead us into doubting the importance of tracker reasons in determining, along with service reasons, the reasonableness of emotions. And here we have the logical space for a more radical discrepancy between reasonable emotion and reasonable action – more radical, I mean, than the discrepancy already shown by the Adams argument. For now we see that the balance of reasons for and against emotions includes a whole class of extra reasons – emotional tracker reasons - that are played out in thought alone. By their nature they do not exist in the service of value and hence do not militate in favour of or against action. Or do they? Is my 'hence' here concealing a nonsequitur? Of course we should concede at once that the existence of emotional tracker reasons has some implications for action. To deny this would be to undermine the whole purpose of our inquiry. A's conformity or nonconformity with emotional tracker reasons, we are arguing, can bear on the reasonableness of A's emotions; the reasonableness of A's emotions can bear, in turn, on the excusability of A's unreasonable actions on the strength of those emotions; and the excusability of A's unreasonable actions can bear, of course, on what ought to be done to or by A in response to those actions. (Should B punish A or reproach A or denounce A or report A or convict A of a crime? Should A atone or reproach himself or turn himself in? And so forth.) But these are not cases in which emotional tracker reasons qualify as reasons for or against actions. They are cases in which *the fact of A's conformity or nonconformity* with emotional tracker reasons qualifies as a reason for or against actions. Our question right now is whether the tracker reasons that help to make A's emotions reasonable can themselves contribute to making A's actions motivated by and manifesting those emotions similarly reasonable, such that he does not need an excuse for those actions. Two possibilities come to mind.

First, couldn't there possibly be tracker reasons, as well as service reasons, for and against actions? Strictly speaking this would require a modification of our definition of a tracker reason, which was characterised as bearing only on thought, but the modification would surely not be too hard to furnish. It would only require us to envisage a kind of practical fittingness, a way in which actions could be apt to an occasion independently of their value. And many have envisaged just such a thing. Many, in particular, have offered tracker-like explanations for reasons of desert, especially desert-based reasons for punishment. These reasons may seem to have at least the following feature in common with tracker reasons: that someone is punished more than they deserve to be for their wrongs is no reason in favour of their committing more wrongs so that their punishment will have been deserved after all. My own view is that the scent of a tracker reason here is deceptive. The explanation of desert-based reasons for action, as of all other reasons for action, must point to a value which is served by conformity with them.[36] The reason why one cannot put an undeserved punishment right by adding new wrongs afterwards is simply that the punishment cannot have been for those very wrongs, since *ex hypothesi* they were not committed at the time when the punishment was inflicted. The value of punishment *qua* deserved depends on its having been punishment for the very wrongs that made it deserved. But even if there are some tracker reasons for action here (maybe they are anomalous in something like the way that service reasons for belief are anomalous?) how would we generalise the point so as to maintain across-the-board parity between tracker reasons for actions and tracker reasons for the emotions that motivate those

[36] I touched on the issue in my introduction to H.L.A. Hart, *Punishment and Responsibility* (2nd ed, Oxford 2008), xiii-xvii. As I indicated there, the fact that one must point to a value which is served by conformity with desert-based reasons doesn't entail that the value must exist independently of the reasons, i.e. it does not prejudge against a deontic interpretation of them.

actions? True, there are emotions, such as indignation and guilt, that seem to track much the same facts about people as do the criteria for determining who deserves to be punished. Perhaps something analogous can be said of gratitude *vis-à-vis* deserved reward. But can anything analogous be said of envy, greed, jealousy, *Schadenfreude*, or shame? My friend's good fortune is enviable. My enemy's ill fortune befits *Schadenfreude*. Probably, as a rule, neither emotion is capable of harbouring enough positive value to be reasonable; yet both are, in the sense we have been considering, apt to the occasion. That is one thing to be said for them. What distinctive action on the strength of them is similarly apt to the occasion? Nothing by way of good or ill-treatment is deserved in either my friend's case or my enemy's, and there seems to be no similar class of superficially tracker-like reasons for action that fills the same space in relation to these emotions that desert-based reasons fill in relation to guilt, indignation, remorse, and similarly 'moralistic' emotions.[37]

Second, even if there are no tracker reasons for actions, aren't there service reasons to manifest one's reasonable emotions in one's actions? If so then the force of tracker reasons for emotions is indirectly reflected in service reasons for acting on the strength of those emotions. Now, we already encountered and endorsed the idea, at the start of section 2, that there can be intrinsic value in the manifestation of reasonable emotions. What we denied was only that this intrinsic value could ever be part of what made the emotion reasonable. Isn't it nevertheless part of what makes the *action* (i.e. the manifestation of the emotion) reasonable? Isn't

[37] For further discussion, see Justin D'Arms and Daniel Jacobson, 'The Moralistic Fallacy: On the "Appropriateness" of Emotions', *Philosophy and Phenomenological Research* 61 (2000), 65 at 82-3. I am using 'moralistic' to designate emotions that respond to tracker reasons that are also moral reasons for certain actions that might be motivated by those emotions. As I said, I doubt whether those moral reasons are also tracker reasons for the actions in question, but my designation is designed to leave this open.

the intrinsic value of manifesting a reasonable emotion a value, in the service of which one has a reason to manifest that emotion? No. The intrinsic value we are talking about here is none other than the value of virtuous action, i.e. action that exhibits and helps to constitute the good character of its agent. As Aristotle says, 'excellence of character is concerned with emotions and actions, in which there can be excess or deficiency or a mean between the two.' He goes on to illustrate:

> [O]ne can be frightened or confident … or feel anger or pity … either more or less than is right, and in both cases wrongly; while to have these feelings at the right time, on the right occasion, towards the right people, for the right purpose and in the right manner, is to feel the best amount of them, which is the mean amount – and the best amount of them is the mark of excellence. Similarly with regard to actions.[38]

Much has been made of this passage, especially about the so-called 'doctrine of the mean'. But I want to pick up a different theme. What Aristotle says of emotion is that its excellence (its contribution to good character) depends upon its reasonableness, not *vice versa*. And with the word 'similarly' he adds, I think correctly, that the same is true of action: action is virtuous because reasonable, not reasonable because virtuous. Its virtuousness, therefore, does not constitute or yield a reason to perform it. It merely summarises and characterises the undefeated reasons that there already are to perform it. When we put this idea about virtuous action together with the companion idea about virtuous emotion we get something like the following.[39] An action is virtuous because it is a reasonable action that manifests a reasonable emotional response (=emotion or lack of it), and it earns its title of virtuous from (*inter alia*) the

[38] *Nicomachean Ethics* 1106^{b}16-24.

[39] For detailed discussion of how the two ideas fit together, see Michael Stocker, *Valuing Emotions* (Cambridge 1996), 169ff.

reasonableness of both the emotional response and the action. In earning that title, it brings a further value to the world, namely the value of the world being inhabited by better people, i.e. people of better character. That value is not, however, a possible reason to perform the action in question.[40] For it is (as we might say) a supervenient value that comes of the agent's correct engagement with *other* reasons, viz. the reasons that make her emotional responses and her actions reasonable in the first place.

This is the grain of truth that I promised to retrieve from the mistaken idea that the fittingness of emotions to their objects is a kind of intrinsic value in them. For the fittingness of emotions can help to lend intrinsic value to them, inasmuch as their fittingness contributes constitutively to their reasonableness and their reasonableness contributes constitutively to the good character of those who possess them. This reveals, incidentally, another helpful way to think about emotional (as well as some other) excuses. They are allocated, we might say, according to standards of character.[41] Think first about the relationship between virtue and justification. Not all justified actions are virtuous. So long as one acts for an undefeated reason one's action is justified whether or not one exhibits a reasonable emotional response in the process (and likewise in the absence of various other constituents of a good character). Justified action is available in principle to virtue-emulators, such as the enkratic. But how about the converse? Can an action exhibit virtue without being justified? Yes it can, but only in an imperfect form. A perfect (model, exemplary, paradigm) exhibition of virtue requires justified action - action for undefeated reasons – but also exhibits a justified emotional response, i.e. an emotional response that one has for an undefeated reason. Where a justified emotional response is exhibited in unjustified action, we have an

[40] It is of course a reason to perform many *other* actions, notably actions of training (or 'habituating') one's character. *Nicomachean Ethics* 1114[a]4-10

[41] This is how I presented them in 'The Gist of Excuses', above note 3.

imperfect variation on virtue. The action lives up to one standard of character but not to another. Yet that is no accident: for it is the justified emotion that motivates the unjustified action. This, it seems to me, is where emotional excuses find their natural home. They represent a certain unlucky, one might even say tragic, imperfection of virtue: the agent was led by one aspect of her virtue, the emotional aspect, to fall short in another aspect of the same virtue, namely its implementation in action.

4. In the thrall of emotion

So how do emotions contribute to the actions that manifest them? What kind of force do they exert? We have seen that emotions are not reasons. Explaining one's action by pointing to the emotion that motivated it is not yet offering a reason for having done it. On the other hand emotionally motivated actions, like all other human actions, are done for (what their agent at the time takes to be) reasons.[42] So in cases of emotionally motivated action the question 'why did you do it?' (e.g. 'why did you run away?') can be answered in two consistent ways: by giving the reason one supposed oneself to have (e.g. 'he was going to kill me if I didn't get out of there') or by citing the motivating emotion that accompanied it (e.g. 'I was terrified').

Reflecting on the relations between explanations of these two types, Hume concluded that one needs to presuppose an independent motivation in order to understand the reason as an explanation for the action. 'He was going to kill me if I didn't get out of there' rationally explains my running away only when

[42] I say 'at the time' because in pathological cases the agent will often be at a loss, after the event, to explain the reason. For example the arachnophobic may be at a loss as to why spiders are frightening. At the time it's simply the fact that there's a spider that the arachnophobic takes to be a reason. If asked why she's running, she cites that very fact: 'There's a spider!'

coupled with 'I was terrified' (or 'I didn't want to die' or 'I love my life' or some other report of motivation).[43] Kant replied that explanations of the two types are, on the contrary, autonomously intelligible. There is a rational type of explanation mentioning reasons alone, and then there is a causal type of explanation mentioning such arational forces as emotions and desires. This does not mean, as Kant pointed out, that emotions cannot provide additional motivation to follow whatever reasons one should anyway follow, and hence cannot have a rationally defensible role in action.[44] It means only that the action is still fully explicable in rational terms – as a following of the reasons - without mentioning the fact that motivational fortification was provided by emotions. Meanwhile the motivational fortification remains, in Kant's words, 'blind in its choice', an arational force that can be exploited by rationality in much the same way that a wind or a tide can be exploited, but which has no built-in dependency upon or answerability to reasons.[45]

This Kantian dualism is often treated as if it were an unnecessarily extreme reaction to the Humean alternative. In one way, however, it still concedes too much to Hume. Kant is right that rational explanations are autonomously intelligible. 'He was going to kill me if I didn't get out of there' is an intelligible rational explanation for my running away. All that needs to be added to connect the reason to my action of running away (and this is generally implied by the explanatory context) is that I acted *for* the reason, which entails that I was aware of its existence

[43] *Treatise of Human Nature* (2nd ed, Oxford 1978), 415.

[44] *The Metaphysics of Morals* (trans Mary Gregor, Cambridge 1996), 204: our natural fellow-feeling can be 'use[d] … as a means to promoting active and rational benevolence' and such use can indeed be a duty. Kant also consistently allows, contrary to a familiar caricature, that taking (Aristotelian) pleasure in acting rationally does not detract from one's action *qua* rational, so long as it does not affect one's reasons for acting:

[45] Ibid, at 216.

at the time. True, I was also terrified at the time, and my terror had me acting for the reason in question when I might not otherwise have done. Yet one does not need to hear about this motivation (or any motivation) in order to understand the reason as an intelligible explanation for the action. The converse, however, does not hold. Contrary to the view shared by Hume and Kant, but in line with the Aristotelian view, one needs to presuppose the reason (or some reason) in order to understand the motivation as an explanation for the action. In the case under discussion, 'I was terrified' requires 'He was going to kill me' to make sense of it as a motivation, and this is because 'He was going to kill me' is both (i) my tracker reason for being terrified and (ii) the service reason to which, in my terror, I give heightened attention as a reason for action, such that the reason has me running away. It is tempting here to say 'exaggerated attention'. But this is another lesson that we already learned: giving heightened attention to one reason as compared with another, in a way that is not explained by their comparative rational force, is not necessarily giving exaggerated attention to that reason. Recall that it is rationally optimal for Jack to care about some things more than others, meaning: to attend to some reasons for acting more than to others in a way that is not a function of their comparative rational force alone.

In this way emotions answer to reasons even as they mediate, and thereby adjust, our attention to those reasons. In cases of particularly strong emotion, meaning emotion that focuses our attention fiercely upon one reason for action, or upon a small number of reasons for action, we may well speak of ourselves as acting 'in the thrall' or 'in the grip' of the emotion, and not merely 'on the strength' of it. Notice that the work done by emotions here is not that of (as it were) applying additional causal impetus to our actions. It is the work of applying extra cognitive emphasis to our reasons, focusing our rational attention. The work that emotions do for us is in this way similar to, but normally an amplification of, the work of belief alone. Both in

our beliefs and in our emotions we capture – as I put it above, 'track' – aspects of the world, including but not limited to facts that constitute reasons for action. And both in our beliefs and in our emotions we attend to some aspects of the world more than to others. For that reason we may sometimes also be said to act in the thrall, or in the grip, of our beliefs. Talk of strong beliefs does not always refer to beliefs that are held with confidence; sometimes it refers to beliefs that are motivationally strong, meaning that they focus our attention intensely on certain reasons for action at the expense of others. And while the explanation of how they do this may be that they are belief-ingredients of strong emotions, it need not be so. It may be that they are simply beliefs about things we care about, on the model of Jack's love of art and his enthusiasm for Chartres.

You may say that the comparison here between belief and emotion is forced. Emotions are partly constituted by beliefs. Am I not exploiting that fact in order to give an excessively cognitive interpretation to the practical force of emotions? It is true that emotions are partly constituted by beliefs, but it is also true, as I explained, that there is more to be captured in the world than what can be captured in beliefs, even in the belief ingredients of emotions. There is more than just the way things are *qua* believable. There are aspects of the world that are only properly captured in complete (wishful and affect-laden) emotions. That is why there are tracker reasons for emotions, tracker reasons that are not conformed to merely by believing. But it is also why emotions are capable of providing an extra degree of rational focus: they focus our attention even more decidedly than a belief alone is normally primed to do upon the facts that constitute tracker reasons for them, and if these facts also constitute or yield service reasons for action, then upon those allied service reasons too. So, for example, Jill in her indignation focuses on her boss's patronising remark, which is the built-in tracker reason for her indignation (i.e. the one which figures in her indignation's secondary belief). But it also constitutes or yields a service reason

for her various attempts to set her boss right (by showing him her mettle and/or by punishing him). The more intense Jill's indignation - the more the component beliefs, wishes and affects of indignation conspire to focus her attention on the patronising remark – the more these allied service reasons for setting her boss right stand out for her, eclipsing other reasons that might otherwise have been ripe for her attention, and hence diverting her from other actions that might otherwise have been on her repertoire for today (whether for good or for ill).

So my use of expressions such as 'in the thrall of' and 'in the grip of' does not point to a lingering affection for, or illicit trade upon, the view of emotions that I officially reject. It is consistent with, and indeed presupposes, that even very powerful emotions are guided by reasons and are accordingly subject to assessment as reasonable or unreasonable. But what about the idea that, in the grip of some emotions, we are 'driven' to act? I agree that talk of emotions 'driving' us invokes a rival view of emotions that I reject.[46] However to the best of my knowledge I have never indulged in such talk. All I have said, I hope, is that, when we act in the grip of some emotions, we are driven to act as we do. It does not follow – nor is it true - that in such cases we are driven to act *by our emotions*. We are driven to act, rather, by the facts to which our emotions make us especially attentive, the same facts which constitute tracker reasons for the emotions in question and service reasons for the associated actions. More generally, it is not emotion that acts on us. It is the world that acts on us by acting on us emotionally. We are irritated by irritating things, shamed by shameful things, amused by amusing things, and so on.[47] The image of the world 'driving' us is merely an extension of this idea of the world acting on us. It is no different from the sense in

[46] See Kahan and Nussbaum, 'Two Conceptions of Emotion in Criminal Law', above note 1, at 280, who regard talk of emotions 'driving' us as characteristic of the 'mechanistic' view of emotions.

[47] See Gordon, *The Structure of Emotions*, above note 13, 110-121.

which we can be driven to act in certain ways even without emotions. It is no different from the sense in which Martin Luther, for example, took himself to be driven by reasons when he supposedly announced '*Hier stehe ich. Ich kann nicht anders. Gott helfe mir. Amen.*' Was he presenting himself as driven by his conscience, such that he could not be held up to any expectation of reasonableness? No. He was presenting himself as driven by those reasons of duty which loomed large in his conscience. And to be driven by reasons, of course, is to invite an assessment of one's being so driven in terms of reasonableness.

Some strong emotions (as well as some strong desires and attitudes) are known alternatively as 'passions'. In order to make sense of this, one need not regard human beings as passive in the face of them. Still less need one hold, with Kant, that 'affect is like an inebriation that makes one sleep, passion is like a lasting madness.'[48] Affect is nothing like an inebriation, passion nothing like a madness.[49] Emotion, more generally, is not a kind of pathology. There certainly are pathological emotions: phobic fear, neurotic guilt, morbid jealousy, and the like. Perhaps, on some occasions, we are not responsible for our actions in the thrall of such pathological emotions. But we have already seen that normal emotions are very different, and far from negating our responsibility, belong to the class of things for which we are responsible (i.e. for which we can be expected to have and to offer rational explanations). How so? Because those who exhibit normal emotions, even when they are in the grip of them at their most powerful, take themselves to have a reason for exhibiting them, and there is a tracker reason for exhibiting them that

[48] The words are from the Mrongovius transcriptions of Kant's 1784–5 *Lectures on Anthropology*, as quoted by Maria Borges in 'Kant on Sympathy and Moral Motives', *Ethic@* 1 (2002), 183.

[49] Although it may not be so outrageous to think of very passionate *love* as like a madness, given that love answers so little to reason. For discussion see Irving Singer, *The Nature of Love: Plato to Luther* (2nd ed, Chicago 1985), 15–16.

matches the one that those who exhibit them take themselves to have. And this is just the start, for many other reasons, service reasons too, bear on their assessment. That being so, a non-pathological emotion is always open to assessment as reasonable or unreasonable. And it is this fact that makes it possible for there to be emotional excuses. For offering an excuse, to reiterate, is a way of asserting one's responsibility, a way of asserting that one answers to, and indeed lives up to, the normal expectation of reasonableness applicable to responsible agents.