A ROSE BY MANY OTHER NAMES:

Emotional Reappraisal, Bootstrapping, Confabulation and Enhancement

Final draft of a paper now published in _Estética das emocões_ ed. Fernanda Gil Costa & Igor Furão, Humus, Lisbon.

> *What's in a name? That which we call a rose, by any other name*
>
> *Would smell as sweet. So Romeo would, were he not Romeo call'd*
>
> *Retain that dear perfection which he owes*
>
> *Without that title.*
>
> (Shakespeare, *Romeo and Juliet,* II-2)

> *[If your name wasn't Ernest,]] I might respect you, Ernest, I might admire your character, but I fear that I should not be able to give you my undivided attention."*
>
> (Oscar Wilde, *The Importance of Being Earnest,* II-3.)

My two epigraphs set the problem with which I shall be concerned in this paper. For Juliet the name of Romeo makes no difference to his essence, or to her experience of him. For Cecily, on the contrary, the name 'Ernest' is indispensably associated with anyone she might love. Nietzsche, in a fascinating fragment of the *Gay Science,* appears to come down on Cecily's side—with a twist, however, for he takes the point to express a fact about consciousness itself. What appears to us to be individual consciousness, Nietzsche contends, evinces a deplorable subservience to the public coin of language. And our reliance on language seems to him to

constitute an impassable obstacle to the development of genuinely individual experience. He elaborates as follows:

> My idea is… that consciousness does not really belong to man's individual existence but rather to his social or herd nature; that, as follows from this, it has developed subtlety only insofar as this is required by social or herd utility. Consequently, given the best will in the world to understand ourselves as individually as possible, … our thoughts are continually ... translated back into the perspective of the herd. (Nietzsche 1974, §354).

The questions I propose to discuss here will be restricted to our emotional experience: Is it indeed the case that when we attempt to articulate emotions, our thoughts get translated back into the perspective of the herd? And how do words, of language in general, affect the quality of emotional experience?

I will start by setting out three framing ideas. I will then speculate about ways in which we could understand the idea that some aspects of our experience are ineffable. I will then turn to the idea that language changes both our understanding and our experience of emotions, and consider how this could be viewed either as a good thing or a bad thing. Insofar as it is described as confabulation, the implication seems to be that it is a bad thing—that there is something false about it. But changes in our emotions could also result from the process we might call "bootstrapping", a metaphor derived from the capacity of computers to set up their own operating system when they are first turned on. Bootstrapping connotes creativity rather than deception. And the difference between those two notions has always called for a degree of subtlety. Our choice of one or the other term to describe the changes brought to our emotional experience by

verbalization, I shall suggest, might be an existential one, reflecting our individual temperament rather than any objective psychological fact.

I Framework Ideas:

(i) the Two Track Mind

I will be taking for granted that the human brain implements what we can call a "two track mind" (de Sousa 2007). This is often referred to as the “dual processing hypothesis”, and is increasingly widely accepted among psychologists and philosophers. There are many versions: Keith Stanovich, in his 2004 book *The Robot’s Rebellion* already listed two dozen versions and there have been several more since. (Stanovich 2004). It posits that we should distinguish between "analytic" and “intuitive" mental processes. The intuitive system includes some innate dispositions, but also procedures that are overlearned and occur as if they were reflexive. They are extremely fast, typically take place without conscious deliberation, and are relatively informationally encapsulated, which is to say that they are not very susceptible to influence by other information. They involve analogue rather than digital representation. The analytic system, by contrast, is typified by explicit calculations or inferences explicitly formulated in language. It is deliberate and involves conscious monitoring. Its drawback is that in comparison with the intuitive it is very slow and somewhat inefficient. Most important, however, is that although language itself has been regarded as a separable functional “module” of the mind, it is topic neutral (Carruthers 2003). Normally, for example, there is no direct communication between your visual and auditory perceptual modules but we can talk about both and about how each relates to the other. Finally the analytic system involves primarily digital representation.

Where do emotions fit into the Two-Track Mind? A moment's thought is enough to realise that emotions straddle the divide. They look like Track One phenomena in that they happen fast and without deliberation. Notoriously it is very difficult to bring someone who is an emotional state to listen to reason or acknowledge contrary evidence. At the same time, however, many emotions have propositional objects, which has prompted a number of philosophers to identify them with judgments or thoughts (Neu 2000; Nussbaum 2003). They are easily triggered by linguistic input. Indeed, as we shall see below, mere exposure to "priming" by single words can have surprisingly powerful effects in triggering emotions even when we have no consciousness of having heard the words in question.

(ii) The centrality of the experiential dimension of experience.

Students of emotion typically stress four aspects or components of emotions: the *physiological*, including changes in the arousal of either one of the antagonistic systems of the Autonomic nervous system; the *motivational*, which results in organized responses, triggered by situation types and paradigm scenarios; the expressive and behavioural, which are crucial to *the social* functions of emotions, and finally the *Experiential*. Many a conscious emotion's felt character is valued for its intrinsic quality, which is orthogonal to the value *signalled* by the emotion regarded as bringing information about the world. Thus horror signals danger and death; sadness signals loss; and yet we go to the movies purposely to enjoy horror and sadness. This dimension of experience, though it has proved of great interest to aestheticians, tends to be neglected in the psychology of emotion. It is with this experiential dimension of emotions that I shall be concerned with here.

(iii) Digital Representation

I have already alluded to digitality as an essential characteristic of language. To understand what this means, it is crucial to remember that the world, as such, is neither analog nor digital. The physics of matter may reveal that there are elementary constituents of space that constitute minimal quanta, or it may turn out that physical space is, like geometrical space, continuous. So it makes sense to ask whether physical space, or some physical quantity, varies discretely or continuously. But that doesn't make it digital or analog: for those terms apply exclusively to systems of representation. The analog-digital distinction is really about representation for the purpose of making copies. In analog copying, accuracy admits of degrees: but however accurately a line is copied, small errors are bound to accumulate in patterns defined by random walks. In digital representation, by contrast, degrees are fixed ahead of the game, in setting out basic atoms of representation. I take it as a given that language is a digital system, as attested, for example, by the very obvious fact that there is nothing corresponding to different degrees of resemblance between, say, 'revelry' and 'rivalry' or between 'bathos' and 'pathos'. In a digital system there are no differences of degree, only differences in kind.

Digital representation is, in fact, an invention of Plato's. It is implicit in Plato's paradoxical theory of resemblance, the Theory of Forms. Our alphabet provides the clearest illustration. Take two tokens that resemble one another, say any two graphics expressing the latter A. Is the resemblance between them a two-term relation? Obviously yes, says common sense. But Plato held that it should really be treated as a three-term relation: what the resemblance between 'a' and *'a'* consists in, he proposed, is the fact that both derive from a third entity, call it 'A', the ''ideal'' which is the depository of the essential nature of both 'a' and *'a'*. There are, in fact, a

thousand different fonts and styles of type, and two tokens of ‘a’ belonging to different fonts might resemble one another less than they might resemble a ‘d’ or perhaps a ‘5’ in some other font. What all instances of the letter ‘a’ have in common is just that they meet whatever norms it is that constitute them as concrete instances of that particular letter.

As everyone knows, computers of the sort many of us now use every day are based on a system of digital representation. What that means is that the voltage changes effected in the computer’s circuitry are regarded, for the purposes of computation, as taking only one or another of a finite, predetermined number of discrete values. From the purely physical point of view, of course, the voltage changes are effectively continuous. This illustrates the importance of distinguishing between the characterization of the physical processes themselves, and the way those characteristics are interpreted when they are set up to be part of a digital representational system.

In the natural world, only two systems are known to be organized as essentially digital representations: the genetic code, and language. This is what they have in common: both are involved in the transmission of indefinitely many copies; and to make that possible, both rely on their capacity for a very high degree of fidelity. Such a high degree of accuracy in replication simply could not be sustained by analog copying. Inside the brain, by contrast, there does not seem to be such an obvious need for highly reliable multiple copying: that fact contains a hint that Nietzsche might be wrong, after all, in supposing that our consciousness is tied to the digital categories to which we are committed when we need to communicate with other members of the common “herd”.

How in particular does this apply to emotions? Of the four central aspects of emotions I have mentioned, it seems plausible that just two, the behavioural and what I have called the social,

give rise to chains of repeated transmission requiring digital representation. Robin Dunbar has suggested that language made *gossip* possible, as a substitute for mutual grooming. As human groups grew larger, gossip enabled our ancestors to keep up with larger numbers of associates (Dunbar 1996). It also demanded larger brains to keep up with the more complex information required, and in turn the cognitive resources unleashed might have given rise to a range of activities that could give rise to chains of transmission requiring accurate copying. Since these might notably include Prediction, Explanation, Generalization, and Gossip, I shall refer to them collectively as PEGGing.

The experiential dimension, by contrast, has in itself no obvious need for accurate multiple copying, and therefore no need for a digital system of representation. This suggests that the temptation to describe emotional experience as "ineffable" *arises from the attempt to translate the analogue system of representation of pre-conceptual experience into the digital system of representation that is embodied in language.* In turn, this hypothesis suggests a modification of the view I quoted from Nietzsche: if what I have said is right, then he was mistaken in thinking that he had identified a property of *consciousness* as such: rather, the property to which he refers stems from the fact that language belongs to the category of digital representation, and arises only when one is attempting to convert experience into communicable linguistic form. That is what necessarily belongs to "the herd", or the social.

How then is individual experience constituted? The answer, it seems plausible to suppose, is that it is constituted by the individuality of each person's life experience, as well as their genetic temperament. And that is true also about the meaning of these public things, the words of the herd. Because each one of us has learned our repertoire of individual words in the course of our

unique individual life, there are subtle differences between any two of our idiolects. (An idiolect is the dialect which belongs exclusively to a single person.) Because my idiolect differs from yours in subtle ways even if we speak the same language, there is only so much that I can communicate to you without distortion. So the lowest common denominator effect which Nietzsche deplores does indeed exist, but it is not due to the fact that we are forced to use the vulgar common categories of the herd. It stems, rather, from the limitation on the range of emotional experiences I can convey that results from the limited number of communicable categories implied by our repertoire of emotion words. If I must rely on a common meaning in the words we both use, there is only so much that I can say to you.

II The ineffable: some conceptual issues

Two interlocking pairs of questions arise about the relation of an emotion to its verbalization. First, we might be enquiring about the fidelity of a verbalization to the emotion it represents. Alternatively, we might be worried about the converse influence: that of the verbalization on the original emotion. Secondly with respect particularly to the latter question, we could ask whether the effect of verbalization on the emotion should be evaluated as a positive or a negative one: does verbalization enhance emotion, or does it blunt or otherwise distort it?

Before approaching these questions, we need to refine our conception of what counts as verbalization. If an emotion is *ineffable*, what is it to *eff* it? Four types of speech acts suggest themselves as potentially relevant: reference, expression, description, and "critical communication". Let us look briefly at each of these.

(i) Reference

The case of reference is straightforward. Whatever is ineffable, if anything is, can certainly be referred to; for if it could not, then the claim that "it" is ineffable would be nonsensical. So whatever else 'ineffable' might mean, 'X is ineffable' cannot possibly mean that X cannot be referred to.

(ii) Expression

What of expression? Verbal expressions of experienced emotion are often equated to non-verbal expressions ('Ah', 'Oh', 'Ouch'…..). These could be adduced as evidence that the emotions thus expressed are ineffable: someone might claim, for example, that words are only used as substitutes for a more adequate form of expression which is purely reflexive and natural. But we might, on the contrary, infer that if cries and moans are adequate expressions of emotion, then the conventional linguistic equivalents of these are also adequate, precisely in virtue of that equivalence. The case of expression, then, is moot.

(iii) Description

Description is the hard knot. There are two ways in which descriptive adequacy might fall short. It might be inferior in informational capacity, considered purely on a quantitative level. Or, on the qualitative level, description might be found wanting because it fails to convey non-conceptual content crucial to the essence of the emotional experience. Let's look at each of these in turn.

It is often said that a picture is worth 1000 words. How many words is an emotion worth? Let us stick with pictures for the moment. How, in the case of pictures or of words, could we make a quantitative assessment of informational capacity?

Steven Pinker has given us a way of making sense of this, albeit in simplified and schematic terms. He begins with very modest assumptions: "Suppose you have ten choices for the word to begin a sentence, ten choices for the second word (yielding a hundred two-word beginnings), ten choices for the third word …. The number of sentences of twenty words or less is 10^{20} : a one followed by twenty zeros, that is, a hundred million trillion, or a hundred times the number of seconds since the birth of the universe." (Pinker 1997, 88) That is an estimate, based on minimal assumptions, on how many twenty-word sentences there could be. On the same principle, 1000 word sentences number 10^{1000} Or choosing from 5000 words of common English, we get 10^{3699} different stories of a thousand words.

Turning to pictures, about one million pixels will fit on a 8"x10" page. If each pixel is either on or off, that's equivalent to some 10^{301030} different pictures. To get some faint sense of the size of such numbers, it's useful to remember that the estimated number of elementary particles in the Universe is of the order of 10^{83}. The numbers we are playing with are therefore superastromical by *thousands* of orders of magnitude: they are so unimaginably large that it actually makes no sense to suppose that a comparison between two of these numbers could be meaningfully experienced. Yet they do represent the number of thousand-word stories, or of one-megapixel pictures, that would fit on one page.

Additional difficulties about measuring such things come from the fact that no single pixel could make as much difference as a single word to the meaning of a whole message. No single pixel in a strictly pictorial pattern amounts to negation, for example. More generally, the amount of information contributed by a single word is not always the same. Although it is true that a picture has a larger content in information in a purely quantitative sense, this takes no account of the

different ways in which the elements of a picture and the elements of a text contribute to its total meaning. A text is constructed of elements chunked into higher level elements: letters make up morphemes, morphemes make up syntactic units of the sort studied by formal grammars (Noun phrases, Verb phrases, qualifying phrases, and so forth); and sentences make up a whole discourse. Focusing on the level of individual words, we will find that the contribution of each element will be a function of the *contrastive set* for a given word in a given context, as well as the *distributive set* of all the contexts in which that word might appear – subject in each case to an appropriate and possibly different contrastive set. Thus in the context 'the — sat on the mat', the word 'cat' has a contrastive set which includes 'dog', 'mouse', etc.; the word 'cat' also has a distributive set that might include the context 'grinning like a Cheshire – ', where its contrastive set might be far more limited. (Jakobson 1970; Ziff 1964). Although there have been attempts to analyze pictures in analogous syntactic terms (see, e.g. Westerhoff 2005), there do not appear to be any normal conventional ways of expressing negation, non-actual possibility, or hypotheticals in purely pictorial terms. That means that whatever the expressive powers of pictures may be, there are advantages to linguistic expressions over pictures in the range of expressive possibilities available to it: for counterfactual possibilities create whole new worlds, which pictures might represent, but which they cannot in themselves label as such. And it is important to remember that counterfactual possibilities are the essence in our emotional life. Indeed, all of our emotions are in part about what might be, at least as much as they are about what is the case; and only language affords us the possibility of specifying *what might be but is not the case*. When we feel fear, or hope, or love, our thoughts are in the grip of possible worlds confronting the actual one.

In sum, there is no compelling reason to think that the informational scope of verbalized emotion is on the whole narrower than that afforded by our visual field. A picture may in a strictly quantitative sense afford more information than an equal amount of space covered with words, but the descriptive space afforded by even a relatively small number of words is so super-astronomically large that the quantitative limitations can hardly be the ground on which emotions are judged to be ineffable.

What of non-conceptual content? Many practitioners of non-verbal arts, performers, dancers, musicians, insist that you cannot convey in words what is conveyed by a dance or by a piece of music. It is certainly true that a description of a visual or musical experience, however skilful, is not a paraphrase of the music or picture; neither will a description ever afford an experience equivalent (in some cognitively or aesthetically relevant sense of that word) to the experience that the work of art is intended to produce. And when you try to articulate it, the content seems to vanish in such a way that you only remember the description that you give: while the experience itself seems to have vanished like a dream. Indeed, although we tend to assume that formulating a description of a visual image is a good way to remember its content, noting remarkable features of the face, for example—long nose, short hair, reddish complexion—there is some fascinating evidence that suggests that under certain conditions verbalization can impede rather than enhance our capacity to recognise a pattern or a face. This phenomenon, called “verbal overshadowing” has been contested; but the experiments that support it can be seen, at least, as casting some doubt on our assumption that encoding our experience in verbal terms invariably solidifies our grasp of it. (Schooler and Engstler-Schooler 1990).

The phenomenon of verbal overshadowing is particularly suggestive in regard to the hypothesis of the Two Track Mind. Much research on the mechanisms of memory suggests that our memories, as well as our perception of the present moment, are largely constructions, if not confabulation. That idea is nicely captured in the title of Gerald Edelman's *Remembered Present* (Edelman 1989). The respective contributions of verbalization and imagery may principally belong respectively to Track Two and Track One. The interaction and occasional interference between the two might explain, together with the difference between cases where verbal overshadowing takes place and cases where it does not, our intuitions about the ineffability of non-conceptual content.

It seems that the idea of non-conceptual content can be shown to have psychological reality even in that domain that seems least likely to exemplify this, namely mathematics. According to Stanislav Dehaene, there seem to be two separate systems governing our sense of the relations of numbers. He demonstrates that human beings as well as animals have a kind of primitive number sense, dealing only with very small numbers, that works independently of our understanding of mathematics based on our knowledge of the rules for manipulating mathematical symbols. The two number systems are controlled by different brain centres and work independently. Track One number sense discriminates one, two, three, and clumps of Many; it is also able to distinguish between larger and smaller collections in a rough and ready way. But only Track Two processes are up to doing any precise manipulations of numbers. (Dehaene 1997). So if even in mathematics we can find a trace of this duality of our processing mechanisms, we can expect that it is indeed a pervasive feature of our mental capacities.

Let us to return now to the proposal that the ineffable aspect of our emotional experience lies in the non-conceptual content—the quale—of perceptual or emotional experience. One very persuasive way of looking at this is contained in these very suggestive lines by the poet Alfred Lord Tennyson:

> *Yet all experience is an arch wherethrough*
> *Gleams that untravelled world whose margin fades*
> *For ever and for ever when I move.* (Alfred Lord Tennyson, *Ulysses*)

Tennyson's reference to a fading margin brings to mind an important point about the nature of perceptual experience in general. In order to describe something we are currently perceiving, we normally focus on a particular point or aspect of the object of perception. In the case of vision there is a literal focal point which allows us to make much finer discriminations than we can make at the periphery of our gaze. As anyone can readily observe, only change or movement are detectable in the peripheral visual field. When we shift our gaze, however, what was in the margin comes into focus; but by the same token something else, perhaps including the original focal point, now passes into the margin and "fades". Now in so far as what is in the margin, while certainly in some (marginal!) sense part of our experience, is very difficult or impossible to describe in detail, it seems that in any instance of visual perception there is in fact an ineffable component. This simply consists in whatever is in the margin; and since there is always something in the margin it follows that there is always something ineffable. There are even surprising experiments in which people can be asked to keep their gaze on a focal point while directing their attention to some peripheral object. Under these conditions, a pattern located at the retinal focal point can be subject to change blindness even though it is, in the literal sense

precisely what the subject is looking at. Attention, it seems, trumps focus. (Mack and Rock 1998).

Extending this observation to the case of emotion requires only an assumption that seems unimpeachable: that all our emotional experience similarly is made up of focal elements, to which the emotion itself might forcefully direct attention, and background feelings that may sometimes seem to modify or even contradict whatever is the focus of attention. And in the case of emotion there is no problem of distinguishing visual focus from attentional focus. So here too we must conclude that our experience always includes something ineffable.

Is this too facile an answer to the question about whether there are ineffable aspects of emotional experience? It may seem to be so because when formulated in this way, the existence of the ineffable is implied by the very structure of our perceptual experience. The advocate of ineffable emotions might deem this to be beside the point, because it is precisely the focal experience that she holds to be ineffable, not the fringe. The present considerations do not touch on this point, and to that extent I still lack both an analysis and a verdict on the question of whether emotional experiences should be deemed to be essentially ineffable.

Furthermore, the fact that verbalization can affect the quality of emotional experience by bringing about a change in focal attention does not by itself say anything about our second main question, whether the change thus wrought in the quality of experience is deemed to be a good thing or a bad thing. I shall return to this point in a moment. What is clear is that the changes that can stem from verbalizing an emotional experience are akin to the phenomenon of reappraisal: sometimes you can reconsider a situation and change your estimate of its character. This is a special case of what we might refer to as the “caption effect”: if I show you a circle with two

straight segments sticking out of it at opposite ends, you will see nothing in particular. But if I tell you it represents a Mexican on a bicycle, you will *see* it. Nothing has changed in the lines you are contemplating, but once you are aware of the caption you interpret them as representations of the Mexican's hat and his bicycle wheels. You don't figure it out: you actually see it. This illustrates our capacity to be guided by a caption in the interpretation that we give of a visual stimulus. Similarly, when a reappraisal is successful we are moved to experience a different emotion by the novel construal of facta that in themselves have not changed.

(iv) Beyond description: "Critical Communication".

So much, then, for the difficulties we encounter when trying to convey a description of an aesthetic or emotional experience. But there are uses of language which are neither simply referential, nor descriptive. This is the case with the language of criticism—not in the sense of assessment of good or bad, but in the sense of 'criticism' that refers to the attempt to guide a viewer's appreciation of a work of art. One might read something like the following in a museum catalogue: "In playful reference to the art-historical tradition, the work exploits the contrast between the jagged contours familiar from cubist themes and softer tones conveying preoccupations typical of post-modern hyperrealism, underscoring post-war optimism with existential angst." This is descriptive, in a way; but when it succeeds it does so, in effect, by means of a kind of extension of the caption effect. As Arnold Isenberg pointed out, the language of criticism aims to give reasons for appreciating a work in a certain way, but is not committed to the universalization of any given formula used in the statement of those reasons. This contrasts with the normal expectations in place when we give reasons for actions as opposed to reasons for appreciation. Isenberg explains this by suggesting that what looks like a reason functions merely

to direct attention; the critic intends thereby to provoke in her interlocutor an experience likely to result from attending to that particular feature of the work. It is then the experience thus generated that constitutes the basis of the critical assessment, rather than the existence of the characteristic designated by the phrase used to direct attention to that feature (Isenberg 1949).

This can sometimes have undesirable effects. The moustachioed Mona Lisa for example, originally due to Salvador Dali, is a good example of artistic blasphemy: once you have seen it, if you should happen to see the original picture—no mean feat these days when it is protected not only by a layer of bullet-proof glass but by a thick troop of Japanese tourists with flashing cameras—you will probably be unable to take it entirely seriously because the moustachioed version will keep popping into your mental gaze. Like blasphemy, this sort of satirical parody changes the focus of our attention in such a way as to disturb and modify the emotional response originally associated with the work, or the divinity, for the purpose precisely of undermining its power. That is why religious authorities not infrequently favour punishing blasphemy with death.

Again there are complications: reappraisals work sometimes, but not always. Consciousness is not actually required for words to have their effect. A well known illustration of this is the placebo effect, which many now think is responsible for a major component of the power of approved pharmaceuticals as well as quack remedies. Placebos testify to the power of words, which essentially belong to Track Two, on the sort of emotional and behavioural dispositions we associate with Track One.

A remarkable set of experiments illustrate this powerful effect of words on the body (Bargh, Chen, and Burrows 1996). In one of these experiments, two groups of subjects read an identical story. But in one group, the story contained a number of words connoting old age, such as

‘white-haired’, ‘Florida’, ‘stooped’, etc. In the other, an otherwise identical story contained words connoting youth and vigour. Subjects were then asked a number of spurious questions about the story; but the actual experiment consisted in observing their progress to the elevator after the experiment had supposedly concluded. The set of words in each group’s story made a significant difference to their posture and speed: the first group walked stooped, and significantly more slowly than the second. Clearly, the subjects had no awareness whatever of either the cause or the effect of this priming process. This exemplified, therefore, the power of words to trigger stereotypes which, in turn, affected subjects' behavioural dispositions. When we think about the power of words to affect our emotions, we should remember the automaticity exemplified in this experiment (and others of a similar nature discussed in the same paper). There is something particularly mysterious about the apparent accessibility of unconscious bodily dispositions to words which we must, presumably, think of as affecting our moods and generally our emotional predispositions.

These experiments represent, as it were, the converse of ineffability. For we say that an experience is ineffable when we feel unable to convey it in words. In this case, however, we are unable to see how the meanings of the words to which the subjects were exposed make their way to their emotional and behavioural dispositions. Their way of being is changed by the causal power of words, but without an intelligible rational mechanism of change of the sort that we are accustomed to attributing to language. Track Two factors seems to be intruding on Track One by borrowing the latter’s non-rational methods or mechanisms.

It is time now to come to the question that I have been repeatedly postponing so far: the question whether the effect of verbalization on emotion is a positive or negative one – a good or a bad

thing. Psychologists use the word ‘valence’ where philosophers tend to use ‘value’. Either word tends to carry the same mistaken presupposition: that there is a single dimension stretching from the extremely unpleasant, through neutrality, to perfect bliss. In fact, however, value is itself multi-dimensional, in the sense that the sources of value are multiple and incommensurable (de Sousa 1974). Something could be positive in value because it's conducive to some adaptive end, or because it constitutes an enrichment of experience, lending it some particular texture or connotation, or perhaps because it is conducive to some particular end. In the light of that, we are unlikely to find a standard answer to the question of whether changes in emotional experience should be seen as enhancing or as distorting.

What we are confronting here is a range of individual differences, owed in great part to ideological commitments or temperamental dispositions. Individuals, in this as in other things, react differently to similar promptings. Remember that each of us has our own idiolect, resulting from the particular circumstances in which we were exposed to the vocabulary we now master, without any memory of that particular learning situation. Similarly, our emotional repertoires have been constructed by individual circumstances in which our specific native dispositions will have dictated subtly different responses (de Sousa 1990). Those responses will in turn have elicited reactions from those around us, in such a way as to reinforce or damp down our propensity to interpret situations and respond to them in those particular ways. Just as the meaning you give to certain words depends on the particular associations they originally came with, so the meaning of our emotional episodes depends on the associations that were present in the circumstances of our upbringing.

The shared language goes with a shared repertoire of emotion words that are used for the purposes of PEGGing: to that extent, mutual understanding does indeed depend on something like Nietzsche's herd consciousness. In Edmond Rostand's drama about Cyrano de Bergerac for example, Roxane's conception of love requires that it be expressed in elaborate and poetic language; the young Christian, however, can only repeat "I love you", and a disappointed Roxane keeps urging him to put it in more interesting ways. But those more interesting ways are not part either of his linguistic or of his emotional repertoire: here the two go together. In order to turn her on he must make it flowery. This is a taste she shares with Cyrano, but it is beyond Christian.

III. Confabulation and bootstrapping.

In the past few decades, it has become increasingly clear that consciousness and memory are both, as I remarked above, constructed as we go. It is well known that people supposedly remember where they were when they heard about Kennedy's assassination or about the crash into the twin towers. But what is remarkable is that the conviction with which people remember such things is actually inversely proportional to the accuracy of their recollection: the more confident you are about what you remember, the less likely you are to be right. We can know this without needing to have access to objective facts, but simply by observing that when asked about the circumstances remembered at a few weeks or months' interval, people give inconsistent answers, not all of which can be correct. (Kensinger 2009). This takes us back to what I have called the anti-Cartesian thesis of the unreliability of our own assessment of first person experience.

Let us now go back to Nietzsche's observation that naming my experience reduces it to the perspective of the herd. His complaint evokes something that most of us have probably had occasion to feel. Having had an experience that seemed to us both subtle and valuable, we hear someone tell a story which is eerily similar and yet somehow repulsive: in their version, the story "gives one the creeps": That isn't it all the way I see it; and yet the words are the same. It felt like a precious moment as I lived it, and yet now that I hear it from those lips it feels vulgar and cheap. One example is provided by Howard Jacobson's *The Act of Love,* in which the narrator, is addicted to the exquisite pleasure of masochistic cuckoldry, learns of someone's fantasy that his wife should wear an ankle chain signalling that she is a "hot wife", a term referring to a practice he has learned of in America. That practice involves the availability of the "hot wife" to other partners with the consent of the husband; and although that description fits his own situation, he finds it appallingly vulgar. (In that case, however, it isn't clear that the difference is indeed ineffable, since the entire novel describes the peculiar masochistic pleasures of his own encouragement of his wife's adultery.)

Conversely, we might have had the experience of thinking how fitting a certain line of poetry is to the present moment. On such occasions we might feel the moment enhanced by that thought, as it lends to us now the poetic flavour of a phrase by Marvell, Donne, or Wyatt. The unpredictability of such moments fits into the general observation that we know more and more about how little we know about ourselves. In particular, as Daniel Gilbert has shown, we are especially inept at predicting the degree of satisfaction or dismay that some future event might cause us. At best, we find ourselves "stumbling on happiness" (Gilbert 2006). Perhaps lowering our expectations about our own predictive success will help; for if our expectations are lower, then our failure to make reliable predictions about ourselves might, depending on our emotional

ideology or on our emotional temperament, save us some distress. On the other hand, it might not: if for any reason an anticipated pleasure doesn't materialize, or merely fails to satisfy, a utilitarian would advise enjoying the anticipation while it lasts: for if the pleasure never arrives, at least the anticipation will have afforded a few welcome hedons. That would seem to constitute a nice enough example of confabulation to deserve the title of bootstrapping. But is there a clear principle of rationality that can justify one attitude or the other? It seems not. For whether we adopt the utilitarian maxim just mentioned as rational opportunism, or deplore it as self-deception, seems to be entirely a matter of individual temperament (see de Sousa 2011, ch. 1).

But if it is all up to arbitrary choice determined by individual temperament, what is the difference between confabulation and bootstrapping? If the imagination constructs something about the past which is not the case then it is self-deception or confabulation. But if, as in some of the cases I've described, you get to construct something which is about the future or the present, and that actually brings on a desirable feeling, then there seems to be no reason to object. If what you believe changes what you feel, that is no reason to deem the resulting feeling to be inauthentic. To cite one last example which seems to favour Cecily over Juliet, consider the following experiment on the pleasure derived from drinking wine. Subjects were given different wines to taste, and they were told that one wine cost $10.00 and another cost $90, though in fact the wines were identical. As might have been expected, subjects rated the expensive wines as having tasted better. But what is remarkable is that this was not an adjustment made after the fact in the light of the price: direct observation of the subjects' brains seemed to show that people were not merely snobbishly saying they enjoyed it more because they didn't want to be seen to be insensitive to quality. Rather, higher priced wines directly caused a higher level of stimulation of the pleasure centres (Plassman et al. 2008). This stunning result is one we should think about

whenever we are tempted to doubt that labels can change emotions. Cecily turns out to have said something deep and correct. The mechanism of this process remains mysterious, but it is undeniable that it exists, and is at one with the result of the Bargh priming experiments, in that the way in which words affect emotional experience is direct and non-deliberate: the Track Two words have a Track One effect.

IV Conclusion

Track One mentality promises at best wordless bliss. Track Two at best promises refinement and elaboration. But whether one regards this as pleasant good fortune or regrettable self-deception seems to depend entirely on what I have been calling one's ideology of emotion.

If that is so, what might such ideologies be linked to? I end with two questions, either of which might well form a good research project for an enterprising doctoral thesis.

First, might a preferences for the first or the second be correlated to certain values on one or more of the Big Five personality dimensions (Agreeableness, Neuroticism, Conscientiousness, Extroversion, and Openness)? (Shiota, Keltner, and John 2006). Openness, Conscientiousness and Neuroticism might turn out to be the most relevant: one might expect, for example, that those who welcome the changes brought about by verbalization of experience might be higher on the Openness scale and lower on the Neuroticism scale. Those who prefer to hold on to the inarticulate and the ineffable might be higher on the Conscientiousness measure.

A second intriguing question for investigation might be whether such preferences are correlated with the ideological beliefs you hold about when and under what conditions you can consider yourself to be most authentically yourself. Those whose existential motto is "You are what you

become” might favour the greatest degree of emotional change that can be managed by redescription and reappraisal. Those, on the other hand, who believe there is greater depth in the converse maxim—“You become only what you are”—might think we should leave our feelings untouched in the primal purity of their ineffable confusion.

References

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology, 71*(2), 230-244.

Carruthers, P. (2003). The mind is a system of modules shaped by natural selection. In C. Hitchcock (Ed.), *Contemporary debates in the philosophy of science.* Oxford: Blackwell.

de Sousa, R. (1974). The good and the true. *Mind, 83,* 534-551.

de Sousa, R. (1990). Emotions, education, and time. *Metaphilosophy, 21,* 434-446.

de Sousa, R. (2007). *Why think? Evolution and the rational mind.* New York: Oxford University Press.

Dehaene, S. (1997). *The number sense.* Oxford: Oxford University Press.

Dunbar, R. I. (1996). *Grooming, gossip, and the evolution of language.* London: Faber and Faber.

Edelman, G. (1989). *The remembered present: A biological theory of consciousness.* New York: Basic Books.

Gilbert, D. T. (2006). *Stumbling on happiness:.* New York: Knopf.

Isenberg, A. (1949). Critical Communication. *Philosophical Review, 54*(4).

Jacobson, H. (2008). *The act of love.* London: Jonathan Cape.

Jakobson, R. (1970). *Main trends in the science of language.* New York: Harper & Row.

Kensinger, E. A. (2009, April). Remembering the details: Effects of emotion. *Emotion Review, 1*(2), 99-113.

Mack, A., & Rock, I. (1998). *Inattentional blindness.* Cambridge: Cambridge University Press.

Neu, J. (2000). *A tear is an intellectual thing.* Oxford; New York: Oxford University Press.

Nietzsche, F. (1974). *The gay science.* (tr. W. Kaufmann). New York: Random House.

Nussbaum, M. (2003). *Upheavals of thought: A theory of the emotions*. Cambridge: Cambridge University Press.

Pinker, S. (1997). *How the mind works.* New York: W. W. Norton.

Plassmann, H., O'Doherty, J., Shiv, B., & Rangel, A. (2008). Marketing actions can modulate neural representations of experienced pleasantness. *Proceedings of the National Academy of Sciences, 105*(3), 1050-1054.

Schooler, J. W., & Engstler-Schooler. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology., 22,* 36-71.

Shiota, M. N., Keltner, D., & John, O. P. (2006). Positive emotion dispositions differentially associated with Big Five personality and attachment style. *Journal of Positive Psychology, 1,* 61-71.

Stanovich, K. (2004). *The robot's rebellion: Finding meaning in the age of Darwin.* Chicago: Chicago University Press.

Westerhoff, J. (2005). Logical relations between pictures. *Journal of Philosophy, 102*(12), 603-623.

Ziff, P. (1964). *Semantic Analysis.* Ithaca: Cornell University Press.