

Are Emotions Valent Embodied Appraisals?

David Pineda (Logos; Universitat de Girona)

The ultimate aim of my contribution is to argue that appraisal theories of emotion (Lazarus 1991, Scherer et al. 2001, Ellsworth and Scherer 2003) fare overall better as accounts of emotions than the perceptual embodied appraisal theory as defended in Prinz 2004, 2007 and 2008. I will try to make my case by relying on three argumentative strategies. First, I will point to certain problems that are inherent to Prinz's view. Second, I will discuss certain crucial issues about emotions and argue that they can be better accommodated by appraisal theories than by Prinz's alternative. Third, I will try to defend appraisal theories from Prinz's main criticisms.

In this paper, I will first present the bare bones of the two theoretical contenders: Prinz's perceptual theory, on the one side, and appraisal theories, on the other (I will be especially brief regarding appraisal theories). Secondly, I will mention, and briefly discuss, some of the arguments I'm going to use corresponding to the three strategies just announced.

1. Prinz's theory.

In his book *Gut Reactions* (Prinz 2004), Jesse Prinz has defended an original perceptual theory of emotion. Part of the originality and interest of the theory lies in the fact that it cleverly integrates elements of so-called cognitive theories of emotions (Solomon 1976, 2003, Nussbaum 2001) together with elements of the James-Lange theory (James 1884), two main theoretical approaches to emotion often thought to be antagonistic. The theory is also compelling because it is argued for after a careful examination both of philosophical considerations and extant empirical studies about emotions.

The first component in Prinz's theory is definitely Jamesian. Prinz argues that (at least in basic cases, more on this later) bodily changes frequently associated with emotions (facial expressions, vocal and musculo-skeletal changes, and changes in the Autonomous Nervous System and the Endocrine System) actually precede emotion rather than following it. James was then right to hold that bodily changes are causes of emotion and not effects thereof. Actually Prinz claims that emotions are perceptions of bodily changes.

There seem to be three main reasons for this Jamesian first conclusion. First, Prinz simply accepts James' "subtraction argument". According to James, if we fancy a strong emotion and abstract from it all feelings of bodily disturbances what we are left with is definitely not an emotion. Prinz reads this thought experiment as showing that the phenomenology of emotion is exhausted by feelings of bodily changes. This is of course a conclusion which squares perfectly well with the view that emotions are perceptions of bodily changes but can hardly be accounted for by cognitive theories according to which an emotion consists of some sort of evaluative judgment or appraisal of the stimulus and as such is a direct cause of the corresponding bodily changes.

Second, he accepts Robert Zajonc's view that emotion and cognition involve separate neuroanatomical structures. In *Gut Reactions*, Prinz mentions Joseph LeDoux's findings about fear (LeDoux 1996) as providing good evidence for this conclusion. LeDoux found out that fear responses to snake-like objects are entirely processed through subcortical regions of the brain. It seems that when the thalamus has the information that the stimulus might be a snake (the thalamus cannot make fine discriminations, the primary visual cortex is required for that task) it sends a signal not just to the primary visual cortex but to the amygdala as well. The amygdala is another very important subcortical structure which is known to orchestrate all by itself the sort of bodily changes typically involved in episodes of fear (Damasio 2010). As the amygdala gets activated the aforementioned changes ensue and usually a typical withdrawal behavior follows quite quickly, before or just when the signal reaches the primary visual cortex. This is why one can sometimes find himself stepping back from a coiled object at the same time one realizes it is not an snake but, say, a house pipe. Now assuming that subcortical brain regions do not implement tasks which require the use of concepts, LeDoux's evidence would then show that some fear responses occur without the mediation of cognitive states such as those that would be required for the sort of appraisals and evaluations postulated by cognitive theories. Of course, LeDoux's evidence, by contrast, is entirely consistent with the Jamesian view that a state of fear is just the perception of the bodily changes orchestrated by the amygdala.

Thirdly, Prinz also endorses the claim, which was already put forward by Karl Lange, that emotions can arise by direct physical induction. The administration of certain drugs seems to have the power to change the emotional state. Consider for instance the ingestion of alcohol and its emotional effects. There seems to be also some evidence to the effect that voluntary acquisition of facial expressions characteristic of the expression of certain emotions actually gives rise to the corresponding emotion (Zajonc et al. 1989). This is again something which a Jamesian theory can perfectly account for. The explanation would be, for instance, that certain drugs have the power to provoke the sort of bodily changes the perception of which is the emotion.

Yet, although for these three reasons Prinz thinks that emotions follow bodily changes and are actually perceptions of these changes, his view is not entirely Jamesian. He claims that emotions are perceptions of bodily changes but they do not represent the bodily changes they perceive. He introduces a subtle distinction between registering and representing to make that crucial point clearer. A mental state, he says, registers that which reliably causes it to be activated. Yet representing is defined drawing on ideas of Dretske and Millikan: a mental state represents that which it has the function to carry information about. Or to put it in the concise terms that Prinz likes to use: a mental state represents that which it is set up to be set off by.

Now according to Prinz, emotions are definitely not set up to carry information about bodily changes. This view, he thinks, cannot adequately explain why emotions were naturally selected as they conferred some sort of survival advantage. He argues that emotions are used to promote certain specific behaviors which are unintelligible if we

assume that they represent bodily changes. For instance, in many cases fear compels us to run away from the eliciting stimulus, but to say that we run away because we feel that certain changes are taking place in our body makes little sense. In an interesting twist in the discussion between cognitivist and Jamesians he claims that emotions represent *core relational themes*. This is, surprisingly enough, a technical term directly borrowed from Richard Lazarus' cognitivist appraisal theory of emotion. Lazarus famously argued that an emotional episode starts when the stimulus is appraised by the organism according to several appraisal dimensions (more in this in the next section, where I introduce appraisal theories). Moreover, an emotion type is individuated by the results of the appraisal process in each of these dimensions. A core relational theme is introduced then by Lazarus as a sort of summary of the results for each appraisal dimension. Therefore, for each emotion type (fear, anger, guilt, etc.) we will find one individuating core relational theme. For instance, in the case of fear, the core relational theme, according to Lazarus, is "facing a danger" (Lazarus 1991, p. 122).¹

As Prinz rightly stresses, core relational themes gloss the bearing of a stimulus on the well-being of the organism. To mention other prominent examples, offense would be the core relational theme of anger and loss that of sadness. If, as Prinz wants, an instance of an emotion type, say a case of fear, represents its core relational theme, then we can understand what kind of survival advantages did emotions confer on our ancestors. Fear would then be a mechanism set up (by evolution) to detect dangers. And we can also understand why emotions compel us to act in certain ways. For instance, it is most reasonable to fly away from an impending danger.

Prinz's central claim is then that emotions represent core relational themes by registering bodily changes. This distinction is further elucidated by Prinz's previous theory about natural kind concepts (Prinz 2002). Consider the concept of dog. This concept applies correctly to something X only if X has a certain complex biological property, a certain genome. Yet, we humans do not have genome detectors. How does our concept then manage to track this biological property? Prinz's answer is that we actually register certain apparent properties (being four-legged, barking, etc.) which are actually caused by the genome in question. To the extent that there is some sort of robust relation between the appearances and the referent we may then develop concepts which track this referent simply by directly registering the relevant appearances. Prinz calls the appearances the nominal content of the concept and the referent the real content. A similar story could be told for the concept water. The nominal content being in this case something as being liquid at certain temperatures, falling from the sky under certain atmospheric conditions, etc., and the real content the property of being H₂O. The idea is then that, for each emotion, the real content is its core relational theme while the nominal content would be the relevant bodily changes. This is why he speaks of emotions as being "embodied appraisals", since they are supposed to represent relations that bear on the well-being of the organism by registering bodily changes.

¹ To be precise, Lazarus distinguishes between anxiety and fright, but we can ignore this complication for the moment. I will return to the issue about types of fear when discussing my arguments.

2. Appraisal theories of emotion.

Appraisal theories constitute one of the main approaches to emotion in contemporary psychology. Although they can be traced back to the ancient Stoics, psychologists often mention the work of Magda Arnold as the seminal source of the view (Arnold 1960). The leading idea is that the emotion undergone by an organism depends crucially on how the organism interprets the stimulus rather than on the nature of the stimulus. This is supposed to explain well-known facts about emotions, namely, that the same stimulus may elicit different emotions in different subjects, or even in the same subject at different times, and that a stimulus may or not elicit an emotion depending on the organism facing it.

Prinz treats appraisal theories in psychology as belonging to the same class, for the purposes of his main argument in *Gut Reactions*, as classical cognitive theories defended by philosophers such as Solomon or Nussbaum. This move assumes that appraisals or interpretations of a stimulus involve some sort of cognitive states and the deployment of axiological concepts. Although Richard Lazarus, probably the most influential appraisal theorist among contemporary psychologists, can perhaps be interpreted as espousing such a view the fact is that some of the current appraisal theorists do not follow him in this respect (see for instance Scherer 2009). This will be of importance later on.

One aspect which is present in most versions of the appraisal theory and is crucially different from such theories as Solomon's and Nussbaum's is that emotions are taken to be causal processes rather than simple states. So, strictly speaking, we should refer to emotional responses as emotional episodes rather than emotional states. According to this view, an emotional episode starts when the organism makes an appraisal of the stimulus.² As happened with Lazarus, current versions of this theory also hold that the appraisal is itself a process which involves the evaluation of the stimulus along a series of parameters or appraisal dimensions. Versions of the general view differ as to exactly which and how many dimensions to count in (see Scherer et al. 2001 for a survey of appraisal theories). By way of illustration, most of them typically include as appraisal dimensions: goal relevance (whether the stimulus, or an aspect of it, bears on some goal or need of the organism); goal conduciveness (whether the stimulus helps to promote some goal or need or it rather obstructs it); coping potential (an estimation of the capacity of the organism to change or modify or conveniently deal in some other way with the stimulus or its relevant consequences, --something which turns out to be crucial

² There is a characteristic hesitation among appraisal theorists on whether to count this appraisal as the initial component of the emotional episode or rather as its triggering cause. For reasons I cannot dwell into here, I think the first option is better than the second. But nothing of consequence follows from this for the purposes of this paper.

when the stimulus has been appraised as obstructive). According to some models, these appraisal dimensions are then processed sequentially. This is all the more reasonable, since some appraisal dimensions seem to require the result of others in order to start on. For instance, an estimation of the coping potential seems to require an output result for the dimension of goal conduciveness as one of its inputs.

The upshot of this appraisal process is the bringing about of a set of characteristic effects. Most models count among them physiological responses (for instance changes in the endocrine system), motor expressions (for instance, facial expressions) and action tendencies (for instance, a tendency to approach or withdraw from the stimulus). This array of effects is supposed to occur more or less at the same time and together constitute the second main stage in the emotional process or episode.

This second stage causes in its turn the third one, which would consist in a representation of most of the elements involved in the previous stages. This representation is thought to be phenomenally conscious.³ This would then in sum amount to the feeling component of the emotional episode, according to these models. The function attributed to this stage ranges from being required for the purposes of communication to being a monitoring device of the whole process which improves its accuracy, efficiency and flexibility. Some models mention also that the modification or suppression of certain emotional behaviors (which might be due to some personal strategic reason or to social norms and pressures) also requires that most elements of the emotional episode be adequately represented by the mind. Finally, some models add also a fourth final stage of verbalization, but we do not need to consider it here.

3. Arguments.

Now that both contenders –Prinz’s embodied appraisals view and appraisal theories— have been summarily presented, let me then mention some of the arguments which build my case that appraisal theories are to be preferred to the embodied appraisals view. In the introduction I mentioned three argumentative strategies. I will therefore start with the first one: problems inherent to Prinz’s view.

3.a Problems of articulation of Prinz’s view.

Perhaps the most important in this score is that there seems to be a serious tension between Prinz’s view of basic emotional responses as having an evolutionary origin and his Jamesian view that bodily changes precede and actually cause emotions. Let me explain.

Let’s go back to the registering / representing distinction and to the central claim that a core relational theme is the real content of an emotion type whereas a certain syndrome of bodily changes is its nominal content. The emotion tracks a core relational theme by

³ According to some models, however, not all elements represented need to be consciously represented (see for instance Scherer 2004). I will not dwell into these niceties here.

registering a certain syndrome of bodily changes. Given how the nominal / real content distinction is explicated this thesis can only be sustained if the bodily changes in question are reliable indicators of the given core relational theme. Compare: the dog concept works only if the appearance properties of dogs through which we track the dog's genome are reliable indicators of the dog's genome. And this is supposed to be so since, in fact, these appearance properties are caused to be instantiated (in normal conditions) by the dog's genome.

So why are the bodily changes reliable indicators of the core relational theme, say, danger, in the case of fear? Prinz's answer is that in the basic cases (which are supposed to be basic because all the rest of cases will be explained in terms of them, more on this later on) there is an evolutionary explanation for this reliable connection:

“Evolution has undoubtedly endowed us with distinctive physiological responses to various situations that our ancestors encountered. The heart is predisposed to race (along with several other physiological responses) when we see looming objects, snakes, crawling insects or shadows at night” (Prinz 2004, p. 69).

So the general idea is that snakes, crawling insects or shadows at night are innate themes for fear.⁴ Evolution has so designed our central nervous system, by way of the natural selection process, that when we perceive any of these themes then a whole syndrome of bodily changes follow. It is evolution then what guarantees that this syndrome is a reliable indicator of any of these themes and a fortiori of danger. But why is such a connection adaptive in the first place? The answer is that this syndrome of changes is known to prepare the organism for an appropriate response to a danger or threat. A racing heart, for instance, enables us to run away. And it is of course highly adaptive to have been provided with a mechanism which enables us to escape from dangers in our (evolutionary) environment.

So far so good, but this does not tell us anything about the natural selection of emotions themselves. Recall that, according to Prinz's view, the emotion is the perception of the syndrome of bodily changes. According to what has been said so far, emotions are idle aspects of the process going from the perception of a theme of fear (say, a snake) to the performance of some behavior adequate to the challenge posed (say, running away). Traits of organisms are naturally selected in virtue of certain effects they have which turn out to be highly adaptive in a given environment. One would like to say (and this is what most people thinking that at least some emotions have an evolutionary origin usually say) that an emotion was selected because it somehow enabled or promoted behaviors which were appropriate for dealing with the sort of challenge posed to the organism by a given stimulus type. For instance, one would hypothesize that fear was selected as a promoter of appropriate responses to dangers. Yet, on Prinz's view the bodily changes which actually enable this sort of adaptive behaviors occur *before* and not *after* the emotion of fear, and they are *causes* of fear rather than *effects* of fear.

⁴ I borrow the terminology here from Ekman (2003).

One may think that this is not a fatal objection to Prinz's general view. One might retort that the previous reasoning only shows that emotions, as perceptions of bodily changes, were not selected as causes of these changes, according to Prinz, but as they had some other effects. Perhaps, just as the appraisal theorist tends to think of the feeling component of the emotional process, Prinz thinks that the effect of emotion which turned out to be adaptive was that of monitoring the bodily changes and therefore of rendering the so called emotional behaviors more flexible and amenable to modification, an effect which would not be in place if the bodily changes were not represented through emotion.

This is not, however, what Prinz seems to have in mind. As I pointed out in the first section, emotions are not supposed to have the function, in the evolutionary sense, to monitor bodily changes and whatever effects derive from this monitoring function. Prinz says that emotions do not represent bodily changes. He says that emotions do not have the function of carrying information about bodily changes; they instead have the function of carrying information about core relational themes. And recall that for Prinz, to carry information about something is just to be reliably caused by this something. So, an emotion like fear is supposed to have the function of being reliably caused by dangers. This is however a bit unfortunate. Again, traits in an organism are not naturally selected for what causes them but actually for some of their causal effects. So fear could not have been selected for being caused by dangers. If one wants to elucidate the notion of representation in terms of functions in the teleological sense and one wants to say that fear represents danger, as definitely Prinz wants to do, the natural move to make is to argue that fear was selected because it promoted behavior which was somehow suitable for dangers. And then we stumble again upon the same problem: the bodily changes enabling the behavior in question are actually the cause of fear, according to Prinz, and not its effect. It looks as if Prinz is at this critical juncture putting the cart before the horses.

This is then the essential tension I mentioned at the beginning of this section. There are two claims which together build up Prinz's characteristic account of emotions: 1) An emotion represents its core relational theme; 2) An emotion is caused by a syndrome of bodily changes which prepare the organism for an adequate response to the instantiation of its core relational theme. The problem is that these two central claims pull in opposite directions if representation is spelled out in terms of functions in the teleological sense, just as Prinz intends to do.

I see also another problem of articulation, which complements with the one just discussed, when Prinz tries to account for non-basic cases of emotion. In Prinz's theory there are two non-basic cases to consider: one brand consists in cases in which a basic emotion like fear is elicited by a stimulus other than an innate theme; the other concerns non-basic emotions. According to Prinz, there is only a limited pool of basic emotions, fear among them, and the rest are derived from the basic ones through two different processes: blending and calibration.

I will make two points about Prinz's treatment of non-basic cases: i) his account of how learnt elicitors of basic emotions arise is doubtful; ii) all non-basic cases are explained as cases in which a syndrome of bodily changes consciously represented is caused by an appraisal of the stimulus. If (ii) is correct, then Prinz's theory looks as something pretty close to what appraisal theorists defend. The crucial difference, of course, is that for Prinz the non-basic cases depend on the basic ones, and actually arise out of them through the sort of mechanisms I'm about to discuss. But as I have just argued, Prinz's account of basic cases involves a tension which renders his account less than compelling.

Let's therefore begin with the first case of non-basiness: elicitation of a basic emotion by a stimulus which is not an innate theme. Consider, to take one of Prinz's examples, being afraid of an exam. It is of course absurd to construe these cases as ones in which evolution has secured some causal link between the mental categorization of a situation as being an exam and the syndrome SF of bodily changes characteristic of fear. Prinz suggestion then is that the thought 'this (the exam) is dangerous (or threatening)' becomes a reliable cause of SF (which in its turn, as usual, causes the state of fear) through a process of associative learning which draws on the basic cases of fear elicited by some of its innate themes.

The problem is that Prinz does not spell out in detail how this learning mechanism is supposed to work. At some point (Prinz 2004, p. 76), he hypothesizes a possible "developmental sequence":

"At some point, while experiencing fear in a darkened room, we entertain the verbally mediated thought that we are facing a dangerous situation. This happens on a number of subsequent occasions. At first, the thought "I'm in danger" is an effect of fear (...) But, through associate learning, that thought becomes a trigger for fear as well. Eventually, the explicit thought "I'm in danger" becomes capable of initiating fear responses in situations that lack the physical features that are predisposed to upset us as a function of biology".

So the general idea is that at our early stages of development we experience states of fear as a result of frequent encounters with fear's innate themes (darkness, snakes, crawling insects). We then develop a concept of danger as a consequence of experiencing all these states. This is supposed to be a concept which "captures the features unifying" these themes (ibid.). As a result of this process, the concept of danger becomes strongly associated with experiences of fear in such a way that the mere application of the concept to a stimulus which is utterly different in nature to any innate theme becomes a triggering cause of an experience of fear.

Now this account makes a number of assumptions about how the concept of danger is acquired which should in any case be empirically confirmed (and the same goes, of course, for the other concepts involved in the rest of basic emotions (such as offense or loss, for instance). On the face of it, a number of questions spring to mind. Is it really required, in order to acquire the concept of danger, that young children experience

relatively frequent cases of fear as a result of encounters with innate themes of fear? What if a young child is lucky enough not to encounter, or at least not frequently enough, such themes as snakes or shadows at night? Will she then not develop a concept of danger? Or will she then acquire a concept of danger which won't become associated with states of fear? We'll have to wait and see whether future research in developmental psychology helps to answer these questions in the way required by Prinz's theory.

Moreover, Prinz claims that danger is introduced as a concept which denotes the unifying features of the innate themes of fear. But, on the face of them, these themes are utterly different regarding their physical nature. Snakes and darkness, for instance, have little in common in this respect. Furthermore, there are perhaps countless ways of grouping together snakes and darkness, but many of them will not group them together with guns or exams (to name a few fear elicitors which are not innate).

Let's now move on to the second type of non-basic cases: Prinz's account of non-basic emotions. Non-basic emotions are supposed to arise out of basic ones by the effect of two different mechanisms: blending and calibration. Blending is just the combination of two basic emotions. For instance, Prinz conjectures that contempt may be a blend of anger and disgust (Prinz 2004, p. 144). This would of course entail, following Prinz's general theory, that contempt consists in the simultaneous perception of the syndrome of bodily changes SA of anger and the syndrome SD of disgust. This is, as far as I know, an empirical consequence that remains to be tested. Be that as it may, Prinz's claims about blending appear very tentative and one has the impression that the mechanism which does more theoretical work is calibration.

Prinz draws again on an idea of Dretske to make clear what is meant here by calibration. According to Dretske, evolved representations can be sometimes put to new uses. Prinz comments on an example outside of the mental realm and derives the intended implication for his theory of emotion:

“Coughing has the evolved function of clearing the throat. But a spy might also use a cough as a secret code in communicating with an accomplice. A spy's cough might represent the fact that the microfilm has been delivered. Likewise, an embodied appraisal that usually represents a demeaning offense (anger) may represent an infidelity (jealousy) when used under the direction of the right judgment. We can recalibrate our embodied appraisals to occur under conditions that are somewhat different than those for which they were initially evolved” (Prinz 2004, p. 99).

In this explanation, it is assumed that anger is a basic emotion which represents offenses and consists in the perception of a given syndrome of bodily changes SA. Then jealousy, a non-basic emotion, is supposed to arise as the bodily changes SA, which are originally calibrated to be caused in general by offenses, get recalibrated so as to be caused also by thoughts of infidelity.

Again this mechanism of calibration raises a number of questions but I will concentrate here on two. Firstly, one may quite naturally ask why these recalibration processes do

occur in the first place. Why are the changes SA recalibrated to be caused by infidelity thoughts? There must be, it seems, some non-accidental connection between the core relational theme represented by anger –offense—and that represented by jealousy – infidelity. This is what Prinz actually suggests. We must recognize, he says, that infidelity involves an offense (see Prinz 2004, p. 148).

The second concern is more serious. Given the answer to our first question there appears to be a serious problem in the way that Prinz individuates non-basic emotions which arise out of a calibration process. If jealousy consists in the perception of the same SA bodily changes as anger and it is also required that the subject judges that infidelity situations are offensive situations, then, given the fact that offense is the core relational theme of anger, one wonders why jealousy is reckoned as a distinct, albeit non-basic, emotion instead of the same old basic emotion of anger.

It looks as if Prinz is individuating here jealousy by a given appraisal of the eliciting situation –namely as one in which some infidelity is involved— and by the fact that this appraisal causes some syndrome of bodily changes and the mental perception of them. Once again, this is exactly what the appraisal theorist does in general. The problem is not merely that recalibrated non-basic emotions turn out to be something quite close to what the appraisal theorist has in mind. The deep problem here is that Prinz seems to be using the appraisal theorist’s way of individuating emotions and disregarding his own.

Prinz tries to evade this problem when he claims that the judgment that one’s lover has been unfaithful need not always be the cause of jealousy:

“Jealousy can be triggered by the judgment that one’s lover has been unfaithful, but it can also be triggered by other judgments, such as the judgment that one’s lover has been staying unusually late at work. Jealousy can even be triggered by perceptual states, such as the smell of an unfamiliar perfume on a lover’s clothes” (Prinz 2004, p. 101).

This is not, however, a good way of evading the problem. Of course, there are countless judgments or perceptions which may give rise to jealousy, as they are countless many others that may give rise to fear, anger or sadness. But of course not anything goes. Only when the organism appraises any of these ways as involving infidelity will jealousy ensue. Otherwise it will not. So this appraisal is unavoidably crucial, it appears, to individuate jealousy.

So the upshot of this section is as follows. The explanation of basic cases of emotion in Prinz’s theory is unstable, since its two characteristic claims –that emotions represent core relational themes and that emotions are caused by bodily changes—pull in opposite directions. On the other hand, the explanation of non-basic cases is problematic on two counts: first, the mechanisms by which non-basic cases are derived are unclear and problematic; second, some of them at least seem to involve ways of individuating emotions which are those of the appraisal theorist and certainly not those which would follow from Prinz’s embodied appraisal account.

3.b Issues better explained by appraisal theories.

Let me now mention two prominent issues about emotions which I think appraisal theories are better equipped to account for than Prinz's theory.

Some of them have to do with what I would like to call the "complexities of the emotional response". In his *Gut Reactions*, Prinz complains that appraisal theorists have mistakenly build the sort of complex properties represented by emotions –danger, offenses, losses, etc.- into the structure of emotions themselves. He is relying here again on another Drestkean point. Consider this example. A simple, unstructured beep emitted from a fuzz buster represents the presence of a police radar. The beep itself is not decomposable in one part meaning "police", another meaning "radar", etc. So the moral is that the complexity of the property represented need not be reflected in the structure of the thing representing it. Appraisal theorists, according to Prinz, have overlooked this possibility in thinking that emotions must be constituted by a complex array of appraisals because of the sheer fact that emotions represent properties of a similar complexity.

But I do not think that this complaint is fair enough. The reason why appraisal theorists take emotions to be highly structured processes is not that they are wrongly assuming that representations must have as much structure as those things they represent. It is rather the fact that the complexities of emotional response and behavior can be hardly explained unless emotional episodes are richly structured. Let me elaborate a bit on that.

Consider once again the case of fear. Not all fear episodes lead, or orient, to the same sort of behavior. When we are afraid of something, we do not always run away from it. In us, as in many other animals, also a fight response is preferred in some cases. And there are still cases in which fear causes a characteristic freezing behavior (which can be conjectured to attempt at camouflage or perhaps to deceive the predator). The appraisal theorist, precisely thanks to the rich structure of the appraisal process which according to her sets off the whole emotional episode, has ways to account for this differential response in different fear episodes. For instance, she can say that a fight, flight or freezing response may critically depend on the result of the appraisal dimension of coping potential.

Prinz is aware of the differential response in different fear episodes. What he says is that there are different types of fear, each one of these types being the perception of a different syndrome of bodily changes. He is clearly assuming that there is a general syndrome of bodily changes which are common or central to every fear episode and then there are some differences which allow us to distinguish between types of fear. This is reasonable, but it shows that Prinz is himself honoring my point: the complexities of emotional response can only be explained by attributing to emotion a more or less rich structure. In his case, this structure is to be discerned in the

relationships between the different syndromes of bodily changes which according to his view individuate emotions.

Yet this may not be sufficient. There are further complexities involved in emotional behavior which are not that easily explained by an account such as Prinz, but can be nicely explained in the framework of appraisal theorists. For instance, it has been suggested that subjects who, for whatever reason, have a low self-esteem will tend to deliver low results in the appraisal dimension of coping potential. As a result of that, appraisal theories predict that such subjects will tend to experience more frequently emotions which are characterized by an appraisal of low coping potential, like sadness (Van Reekum and Scherer 1997). It has also been argued that the optimism-pessimism personality dimension provides predictable systematic biases for the appraisal dimension of goal conduciveness (Scheier & Carver 1985). It's hard to see how the incidence of personality traits, or temporary personal conditions, in the emotional life of an organism can be accounted for on Prinz's account.

The second issue I want to bring up has to do with the connection between emotion and motivation. Embodied appraisals are supposed to be states which track core relational themes by registering changes in the body. They are therefore doxastic states with a mind-to-world direction of fit (Searle 1983). They are states characterized by their tracking, detecting, representing states of the external world and conditions of the body. Yet emotions are commonly thought to be strong motivators for action. Emotions play a crucial role in motivation. This is surely why emotion is a chief topic in psychology since motivation seems to be the key to understand behavior. If emotions were embodied appraisals the role of emotions in motivation would be overshadowed.

Prinz is perfectly aware of this and his final proposal is that emotions are not mere embodied appraisals but embodied appraisals coupled with a valence marker. A valence marker is taken to be a mental state with an imperative content. There are supposed to be two valence markers with the imperative content "More of this!" or "No more of this" corresponding to a positive or negative valence respectively. All emotions have either positive or negative valence (for instance, sadness, anger or fear have negative valence; pride or joy are positive) and Prinz claims that this is so because all emotions include either a positive or a negative marker.⁵ Thus Prinz's final and complete theory is that an emotion is a valent embodied appraisal.

There are many points of interest in this construal. For instance, the idea of mental states with an imperative content is particularly intriguing, but I must leave discussion of this for a better occasion. The point I want to make now is that Prinz's construal of emotion as a conjunction of two utterly distinct types of mental states faces the usual problem with such conjunctive theories: nothing in principle prevents one of the conjuncts to be instantiated in the absence of the other. There might be cases, if Prinz's theory were right, in which the sort of embodied appraisal belonging to fear or sadness

⁵ There might be emotions with a mixed valence, for instance nostalgia, but we can put these cases aside for the purposes of this paper.

is instantiated in the absence of its negative marker or, worse still, together with the positive marker “More of this!”. One can perhaps make sense of exceptional cases in which sadness or fear enjoy a positive valence, I very much doubt this, but let’s assume that this is so. The problem, however, is that since embodied appraisals and valence markers are completely different mental states, with anything interesting in common (the former having indicative content and the latter imperative content), I can’t see nothing in Prinz’s theory that prevents this sort of cases from being the norm rather than the exception. And this is surely an implausible consequence.

Appraisal theories, I think, fare again better in this regard. Appraisal theorists usually explain valence in terms of the net result of some appraisal dimension. For instance, it has been explained as the result of the appraisal dimension of “goal conduciveness”, positive valence arising out of an estimation of goal congruence and negative valence out of an estimation of goal obstruction (Lazarus 1991). Other theorists have instead suggested an appraisal dimension of “intrinsic pleasantness”, one of the first dimensions to be processed, which would estimate whether the stimulus is expected to produce pleasure or pain, broadly construed (Scherer 2001).

Prinz’s dismisses this sort of explanations as being too “overly cognitive” (Prinz 2004, p. 168), but his judgment seems to be based on the premise that appraisals always involve sophisticated, cognitively demanding mental states, a premise simply unshared by most appraisal theorists (more on this in the next section).

The point I want to stress is that appraisal theories seem to have the resources to explain valence as one natural consequence of the very appraisal process and therefore it can be sustained that the valence of a mental episode is inherent to it. This I think is a more promising account.

3.c Reply to Prinz’s Jamesian arguments.

So far I have raised a number of objections to Prinz’s account of emotion and I have mentioned some crucial issues which I think are better explained by appraisal theories. Yet, as I said at the beginning of this paper, Prinz has offered three main reasons for his Jamesian conclusion that bodily changes precede and actually cause emotions rather than the other way around. This runs not only against common sense, as James knew perfectly well when he claimed that his view was contrary to the “natural way of thinking”, but it is definitely incompatible with the central claim of appraisal theories. For, according to these theories, bodily changes involved in an emotional episode are mainly efferent effects produced by the result of different appraisal dimensions. I will then conclude by examining Prinz’s three reasons. My conclusion will be that they do not overall constitute a case against appraisal theories, although some of them may point at matters of concern which need further development and refinement.

The first reason is James’ subtraction argument according to which the phenomenology of an emotional episode is exhausted by feelings of bodily changes. This is a crucial premise for Prinz’s theory. If the phenomenology of an emotional episode were not

exhausted by feelings of bodily changes, then it couldn't be sustained that an emotion is (leaving aside valence) a perception of bodily changes. Yet I expect many defendants of the existence of cognitive phenomenology to be unimpressed by James' argument. These thinkers claim that there is something it is like to think that something is the case (Pitt 2004). Some of them have also carried over this general view into the specific case of emotion. Thus, according to some thinkers, the phenomenology of a sadness episode includes not just feelings of bodily changes but also, among other things, an experience of loss (Goldie 2002, Kriegel 2011).

The claims that there exists in general a cognitive phenomenology and that the phenomenology of an emotional episode includes a cognitive part are controversial and I myself would like to suspend my judgment about them. In any case, I expect Prinz to reject them, specially the second one, which is overtly incompatible with his view.

Even so, and that is the main point I want to make about this first reason, appraisal theories, as they are commonly conceived, have I think little to fear from it. For even if we were to endorse, by accepting James' subtraction argument, that the phenomenology of emotion is exhausted by feelings of bodily changes, we could still hold that this is the phenomenological content of the feeling component of an emotional episode, the third stage of the emotional process devised by appraisal theorists which I discussed in the second section. There is nothing, I think, that prevents appraisal theorists from claiming that an emotional episode becomes phenomenally conscious only when it reaches its third stage, the "feeling component", and that when it does so what we feel are just these bodily changes.

The point is that Prinz seems to be arguing here abductively, using an inference to the best explanation. His argument seems to be this: first, James is right, the phenomenology of emotion is exhausted by sensations of bodily changes; second, the only explanation of this is that bodily changes come first and emotions are caused by them. But the second premise can be challenged. An explanation of James' finding might also be that an emotional episode is a process which starts with an appraisal then it is followed by a syndrome of bodily changes and finally reaches a state of phenomenal consciousness of these changes.

Let's then move on to the second Jamesian reason. Prinz backs Robert Zajonc in his dispute with Richard Lazarus and endorses the claim that emotion and cognition involve two distinct neuroanatomical structures. Of course, this claim is more or less plausible depending on how the elusive notion of cognition is spelled out. Prinz has his own way of doing this which I will not discuss here, but the intended view is that basic cases of emotion like for instance those unveiled by LeDoux's research with fear responses to snake-like objects do not involve cognitive states at all. Prinz, as Zajonc, draws from this the conclusion that they do not involve appraisals of any kind.

This is of course something that appraisal theorists simply deny, and expectedly so since they of course attribute to other animals and new-born infants the capacity to emote. And furthermore they think that the human emotive system has an evolutionary

origin, it was naturally selected as a way of discerning in the environment matters of importance for the organism and of helping it to deal with them in appropriate ways. This being the case, the existence of pre-wired emotional responses, or predispositions to respond emotively to certain stimuli allegedly prominent in our evolutionary past, is to be expected.

Consequently, most models in the appraisal approach have it that some appraisal dimensions, especially those processed first, are frequently run in an automated fashion and in many cases do not involve any cognitive states in the sense of 'cognitive' which bothers Zajonc or Prinz. In fact, some theorists argue that the last appraisal dimensions to be processed in an emotional episode, which would involve an estimation of the bearing of the stimulus or its consequences with respect to social norms or the subject's ego ideal, are in principle only present in humans (and to a limited extent perhaps also in some other primates) and require considerable cognitive effort.

The general idea is that the human emotive system is built upon an evolutionary basis and the cognitive sophistication of the human brain allows it to be more subtle in the sort of appraisals performed and more flexible in the sort of responses given as output. This of course only adds adaptability and efficacy to the whole system. The main idea of appraisal theories is that the emotional response is determined by the subjective evaluation of the stimulus along several appraisal dimensions. This may be hard-wired for the most part, that is to say, the causal connection between a certain result of a certain appraisal dimension and its efferent effects may be hard-wired. This would explain why, for instance, an episode of fear elicited by a snake and an episode of fear elicited by a job interview may largely involve the same efferent responses. But, of course, the processing of each appraisal dimension, the sort of process leading to an output result for a particular dimension, may be more or less cognitively mediated, depending on the cases and the dimensions.

Of course, the interplay of pre-wired and learnt elements in the delivery of the emotional response is at present still little understood. Some theorists claim that more cognitively demanding processes are only called into action when the simpler automated processes cannot solve by themselves the problem posed by the stimulus (Scherer 2001). I'm not persuaded by this view since in many cases a stimulus will typically be felt as posing a "problem" only when it is appraised in certain ways with the help of higher cognitive states. For instance, as far as automated processes are concerned, there is nothing wrong with a gun as opposed, say, to a snake. No problem then to solve and no need to worry about guns as far as these pre-wired mechanisms are concerned. It seems clear that things do not work this way.

Most appraisal models include a dimension of "urgency" which would estimate the need of a fast response to the stimulus. It might then be that when an innate theme for a given emotion is perceived, say a snake in the case of fear, this appraisal dimension delivers the highest degree of output as a result of a fully automated process and that our emotive system is pre-wired in such a way that when the urgency dimension delivers an

output response at its highest degree, or above a given threshold, it causes an appropriate behavior without waiting for the intervention of more sophisticated cognitive processes. This is only a bald conjecture on my part with no empirical support as far as I know, but it is only intended to show that there is room within the framework of appraisal theories to accommodate such as empirical results as LeDoux's.

It remains only to be considered the last reason afforded by Prinz for thinking that bodily changes precede emotion. This is the alleged existence of cases of direct physical induction of emotion. Administering certain drugs or even adopting certain characteristic facial expressions is told to provoke an emotional response. Of course this is on the face of it a fact which seems more easily explained by Prinz's theory than by the appraisal theory. Indeed a plausible explanation of what goes on in those cases is that drugs may cause the bodily changes the registration of which constitutes the emotion according to Prinz. It seems on the other hand harder to defend that drugs may affect the sort of appraisals which according to appraisal theorists give rise to emotions.

It is interesting to note that most appraisal theorists claim that this sort of cases is not covered by the theory. This may be one of the reasons why psychologists tend to speak of appraisals as being the usual cause of an emotional episode rather than its first and triggering component. The idea is that they are the usual way in which emotional episodes arise, but there are also other ways and direct physical induction would be one of them.⁶

I think on the contrary that this reaction may be too quick and that it actually underestimates the resources available to appraisal theories. I will conclude then suggesting some ways in which cases of physical induction can be accounted for by them. As was the case with the previous objection, my aim here is not primarily to argue for any of these ways but rather to show that these cases need not be seen as posing an insurmountable problem to the appraisal approach.

One consideration is this. We have seen that feelings of bodily changes constitute the phenomenology of an emotional episode at least to a great extent. Moreover, it is reasonable to assume that the part of an emotional episode which is amenable to verbal report and conscious recognition is precisely that part which is phenomenally conscious. There can be little wonder then if in cases of physical induction a subject reports feeling something very similar to standard cases of emotional episodes, if not plainly the same. This being said, it is quite a different matter if what is felt in these cases is the genuine thing. Some elements of the emotional episode are probably present—a syndrome of bodily changes, a conscious representation of them—but other crucial ones—the appraisal process—are missing. This is perhaps why drinking alcohol in order to change sadness feelings into joy feelings is not in the long run satisfying at all.

A second consideration allows us to go a bit further. Little is yet known about mental architecture as implemented by the human brain. But some of the best known processes

⁶ Another case would be listening to instrumental music (Scherer 2001). Some philosophers who favor the appraisal approach to emotions also hold this "pessimistic" view, see for instance Scarantino (2010).

show that the brain uses recurrent networks. In recurrent networks information does not travel in one direction only, always forward till the last stage of processing is reached. Instead the output of a certain stage of the process feeds back and influences the output of a previous stage which is processed again and again until the general process stops (Damasio 2010, chapter 3). Some appraisal theorists favor this sort of architecture for appraisal processes involved in emotion and even for the whole emotional episode. According to this, appraisal dimensions which are processed earlier receive input from the results obtained by the processing of later dimensions. The stimulus is appraised and reappraised along the different dimensions and the results obtained in one dimension are influenced by those obtained in the others until the process stops, and this probably occurs when results remain stable and unaltered for some period of time or when the urgency dimension recommends action. Likewise, it is thought that the connections linking the processing of appraisal dimensions with efferent effects are also two-way, with signals travelling forward and backward (Scherer et al. 2001). If this is indeed so then it can be sustained that certain efferent effects brought about by direct physical induction may affect the appraisal process inherent to emotion in such a way that crucial brain sites for emotion may reach a pattern of activation closely resembling that of a standard emotional episode.

REFERENCES.

- Arnold, M. (1960): *Emotion and Personality*. Columbia University Press.
- Damasio, A. (2010): *Self Comes to Mind. Constructing the Conscious Brain*. William Heinemann.
- Ekman, P. (2003): *Emotions Revealed. Recognizing Faces and Feelings to Improve Communication and Emotional Life*. Henry Holt and Company.
- Ellsworth, P. & Scherer, K.R. (2003): "Appraisal Processes in Emotion", in Davidson, R., Scherer, K., & Goldsmith H. (eds.): *Handbook of Affective Sciences*. Oxford U.P., pp. 572-595.
- Goldie, P. (2002). "Emotions, Feelings and Intentionality", *Phenomenology and the Cognitive Sciences* 1: 235-254.
- James, W. (1884): "What is an Emotion?", *Mind* 9: 188-205.
- Kriegel, U. (2011): "Towards a New Feeling Theory of Emotion", *European Journal of Philosophy* (forthcoming).
- Lazarus, R. (1991): *Emotion and Adaptation*. Oxford U.P.
- LeDoux, J. (1996): *The Emotional Brain*. Simon and Schuster.

- Nussbaum, M. (2001): *Upheavals of Thought. The Intelligence of the Emotions*. Cambridge U.P.
- Pitt, D. (2004): "The Phenomenology of Cognition; or What is it Like to Think that P", *Philosophy and Phenomenological Research* 69: 1-36.
- Prinz, J. (2002): *Furnishing the Mind: Concepts and their Perceptual Basis*. MIT Press.
- Prinz, J. (2004): *Gut Reactions. A Perceptual Theory of Emotion*. Oxford U.P.
- Prinz, J. (2007): *The Emotional Construction of Morals*. Oxford U.P.
- Prinz, J. (2008): "Précis of Gut Reactions", *Philosophy and Phenomenological Research* 76/3: 707-711.
- Scarantino, A. (2010): "Insights and Blindspots of the Cognitivist Theory of Emotions", *British Journal for the Philosophy of Science* 61: 729-768.
- Searle, J. (1983): *Intentionality*. Cambridge U.P.
- Scheier, M.F. & Carver, C.S. (1985): "Optimism, Coping, and Health: Assessment and Implications of Generalized Outcome Expectancies", *Health Psychology* 4. 219-247.
- Scherer, K.M. (2001): "Appraisal Considered as a Process of Multilevel Sequential Checking", in Scherer et al. (2001), pp. 92-120.
- Scherer, K. R. (2004): "Feelings Integrate the Central Representation of Appraisal-Driven Response Organization in Emotion", in Manstead, A., Frijda, N., & Fischer, A. (eds.): *Feelings and Emotions: The Amsterdam Symposium*. Cambridge U.P., pp. 136-157.
- Scherer, K.R. (2009): "The Dynamic Architecture of Emotion: Evidence for the Component Process Model", *Cognition and Emotion* 23/7: 1307-1351.
- Scherer, K. R., Schorr, A. & Johnstone, T. (eds.) (2001): *Appraisal Processes in Emotion: Theory, Methods, Research*. Oxford U.P.
- Solomon, R.C. (1976): *The Passions*. Doubleday.
- Solomon, R.C. (2003): *Not Passion's Slave*. Oxford U.P.
- Van Reekum, C.M. & Scherer, K.M. (1997): "Levels of Processing in Emotion-Antecedent Appraisal", in Matthews, G. (ed.): *Cognitive Science Perspectives on Personality and Emotion*. Elsevier Science, pp. 259-300.
- Zajonc, R., Murphy, S., Inglehart, M. (1989): "Feeling and Facial Efference: Implications of the Vascular Theory of Emotion", *Psychological Review* 96: 395-416.

