

## *Do Emotions Represent Values by Registering Bodily Changes?*

*Eva-Maria Düringer*

---

**Abstract:** This paper outlines Jesse Prinz's theory that emotions represent values by registering bodily changes, discusses two objections, and concludes that Prinz's theory stands in need of modification: while emotions do represent values, they do not do so in the first place by registering bodily changes, but by processing information about how things we care about fare in the world. The function of bodily changes is primarily to motivate and prepare us for action.

**Keywords:** emotion, value, intentionality, teleosemantics.

---

Enquiries into the intentionality of emotions seem to take place on the following spectrum: at one extreme is the view that emotions are evaluations of objects; at the other is the view that they are mere registrations of bodily changes. Both sorts of account seem wrong: emotions are not mere evaluations of objects, as they do involve a bodily dimension. When I'm angry, I haven't simply made the judgement that a remark was offensive, but I *feel* enraged: my cheeks are flushed and my heart beats much more quickly. Also, emotions are not mere registrations of bodily changes, as they do involve an evaluative dimension. When I'm angry, my anger does not consist in flushed cheeks and an increased heartbeat alone, but also in the evaluation of a remark as offensive. Though neither of the sorts of position to be found at the extremes of the spectrum seem entirely correct, there is room for many views in between, which try to combine the insights of both ends without having to commit to the disadvantages of either.

Perception models of emotions are one shape these middle ground views can take. Perception models all take emotions to be perceptive states, but they differ according to the object that they take to be per-

ceived by emotions. One perception model claims that emotions perceive values, just like our sensory modes of perception perceive secondary qualities.<sup>1</sup> Another perception model, put forward by Jesse Prinz in his 2004 book *Gut Reactions*, is more indirect than this: emotions do not perceive values directly – rather, it is by perceiving patterns of bodily changes that we come to represent something as being of a certain value. When we are angry, for example, it is by registering that our heart rate has gone up, and that our cheeks are flushed, etc., that we come to represent the preceding remark or action as offensive. If successful, this perception model manages to combine the bodily dimension and the evaluative dimension of emotions: bodily changes play a crucial instrumental part, while values are the actual objects that are being detected.

In what follows, I want to examine Prinz's perception model. In *Gut Reactions* he unfortunately spends only three pages on explaining how it is that we come to represent something as good or bad by registering bodily patterns. But he draws heavily on an earlier paper of his, 'The Duality of Content' from 2002. In the first section, I will outline his view using the insights of both works, in order to get a good grasp of what is at stake. In the sections following I will discuss two major objections to Prinz's model: the first, put forward by David Pugmire, criticises Prinz's use of the notion of a real essence of values. I will argue that Prinz can meet this objection. The second objection aims at what is, to my mind, an overemphasis of the representative function of bodily changes. I will argue that the function of the bodily changes involved in an emotion is first and foremost to motivate us to act. If this argument is right, then this leaves Prinz with the option to interpret the function of bodily changes as twofold: they represent *and* direct. I will show that a model that takes emotions as pushmi-pullyu representations allows for such an interpretation, but also gives rise to new problems. My overall conclusion will be that we should, along with Prinz, take values to be core relational themes, but contra Prinz, take it that our ordinary cognitive facilities manage their representation, whereas the bodily changes involved prepare us for the appropriate action to take.

<sup>1</sup> See, e.g., de Sousa (1987), McDowell (1998), Johnston (2001), Deonna (2006), Döring (2007).

## 1. The idea: representing values by registering bodily changes

The initial question that might arise when considering the view that emotions represent values by registering bodily changes is how it is possible that one can get from registering one thing to representing something completely different. How is it possible, for example, that I get from registering that my cheeks are flushed and my heartbeat has quickened to representing the cause of these bodily phenomena as offensive? Prinz answers this question by drawing an analogy. He compares representing values by registering bodily changes with the way in which we track members of natural kinds. On this picture, representing a remark as offensive by registering flushes and an increased heartbeat is somewhat similar to representing a creature as a dog by registering its furriness and the fact that it barks and wags its tail. I will first explain how Prinz believes we track natural kinds, and then how he thinks this applies to emotions and values.

When we talk about natural kinds, we believe, however vaguely, that there is some essence that is shared by all the members of a group—an essence in virtue of which they belong to this group, that defines this group. All the substances we call water consist of  $H_2O$ , all the objects we call gold contain AU, and all the creatures we call dogs possess the dog genome. And a substance is water in virtue of consisting of  $H_2O$ , a metal is gold in virtue of containing AU, and a creature is a dog in virtue of having the dog genome. These essences, which make an object the kind of thing it is, seem hard to track. After all, we don't see  $H_2O$ , AU or dog genomes, and yet we constantly and correctly classify things as water, gold or dogs. How do we do this?

In order to explain this, Prinz draws on the Lockean distinction between real and nominal essences. Water has both a real and a nominal essence: its real essence is that it consists of  $H_2O$ ; its nominal essence is that it is liquid, transparent, and tastes of not much. Nominal essences are thus *appearance* properties: water feels liquid, looks transparent, and tastes of not much. None of these properties are necessary for something to be water. Something could be solid, blue, and taste of strawberry, and yet still be water. Real essences, on the other hand are defining properties. They are what something has to have in order to be the kind of

thing it is. For something to be water, as we have said, it might be solid, blue and taste of strawberry, but it must consist of  $H_2O$ . Thus, the possession of nominal essences, that is, the characteristic appearance properties of members of a certain natural kind, is neither necessary nor sufficient for something to be a member of that natural kind. The possession of the real essence, that is, the defining set of properties of that natural kind, is on the other hand both necessary and sufficient for something to be a member of that natural kind.

The important point is that Prinz claims that we track real essences by registering nominal essences. We come to represent something as water, because we realize that it is liquid, transparent, and doesn't taste of much. We come to represent something as gold, because we realize that it is bright, yellow, and malleable. We come to represent something as a dog, because we realize that it is furry, barks, and wags its tail. Even though we know that none of these appearance properties is necessary or sufficient for something to be water, gold, or a dog, it is by registering their nominal essences that we track them. There does not seem to be a direct access to real essences. Locke, in fact, believed them to be unknowable. And even though Prinz does not share this scepticism, it is indeed not the case that we ordinarily identify something as water because we somehow know it contains  $H_2O$ , but because we observe the appearance properties characteristic of water. Similarly, we do not ordinarily identify something as a dog because we know it has a certain genome, but because we observe that it barks and is furry. Thus, Prinz claims, we ordinarily track real essences by registering nominal essences. Of course this thesis raises many questions. If we don't ordinarily have access to real essences directly, when can we ever say for sure that what we represent as water is actually water, and not some other substance that happens to be liquid, transparent, and without flavour? Or when can we say for sure that it is not twin earth water? In other words, when can we justifiably claim that a mental representation of an X refers to a real X, where it is impossible to perceive the real essence of X directly? The following diagram should help to illustrate Prinz's view, and the problem that arises:

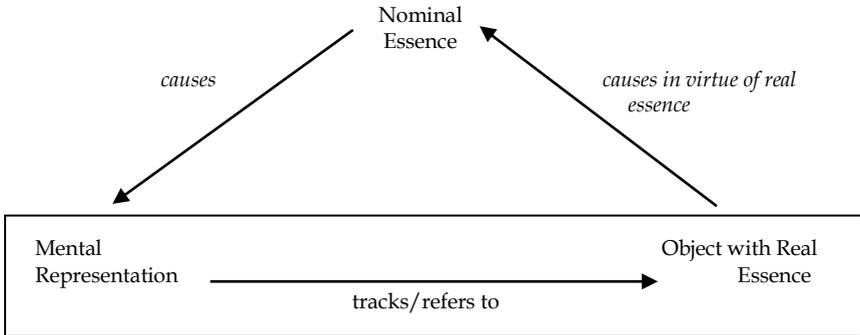


Figure 1: Tracking real essences by registering nominal essences.

When can we say that the box actually holds? Prinz develops two conditions that have to hold between a mental representation and its object in order for the mental representation to be really referring to a member of a natural kind. One of these conditions is the etiological condition, which helps to handle twin earth cases. Since this will not be relevant when it comes to emotions and values, I will not go into it here. The other condition, however, will be important. It is the condition of nomological covariance (NC), which has two parts. (NC) states that the box holds if and only if:

(NC1) objects with a particular real essence causally covary with instances of particular mental representations,

and

(NC2) they do so in virtue of containing this particular real essence.

In other words, not only do particular mental representations and objects with a particular real essence have to causally covary, but the objects must have caused the mental representations in virtue of containing the particular real essence.

The difference between (NC1) and (NC2) is easily illustrated by pointing to figure 1: (NC1) ensures the basic causal covariance between particular mental representation and object with particular real essence. For a representation of X to refer to an object with the real essence of X there has to be a lawlike causal chain that links the two. But this link might not be enough. For there might be a substance that consists of H<sub>2</sub>O

that reliably causes WATER representations, but the causal chain that in fact links the two is that the substance is green, and I mistakenly have learned to identify green liquids with water. Here we have a particular representation and an object with a particular real essence that causally covary, but they do not covary in virtue of the object causing that representation *in virtue of* this real essence. For this reason we need (NC2). (NC2) ensures that the causal chain that links the particular representation and with the object of the particular real essence is indeed a causal chain that goes from the real essence to representation, rather than from some other property of the object:

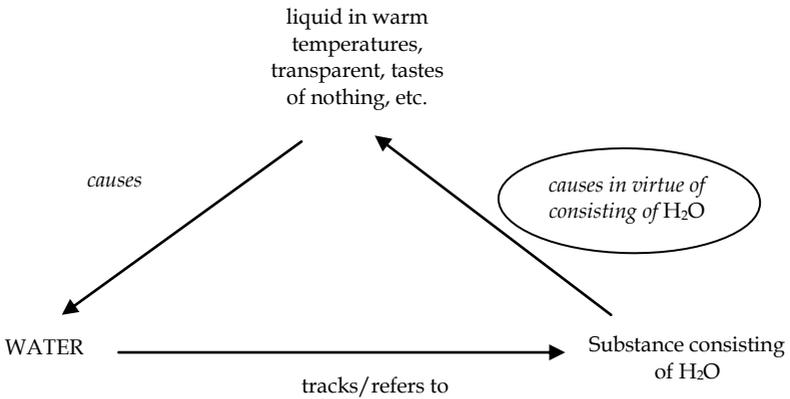


Figure 2: (NC2) explained.

The oval box illustrates what (NC2) demands. For a representation of an object to be a representation of water, the object must have caused this representation in virtue of consisting of H<sub>2</sub>O. If we come to have a WATER representation after having registered that there is something that is liquid, transparent and tastes of nothing, then this representation only refers to real water if the properties ‘liquid, transparent and tastes of nothing’ have been brought about by the property H<sub>2</sub>O.

In the case of the green substance that happens to be water and reliably causes me to token WATER representations, (NC2) is not fulfilled. The nominal essence ‘green’ is not caused by the substance’s consisting of H<sub>2</sub>O. We can only say that a representation of X refers to an object with the real essence of X if the nominal essence by which we come to have this representation has been caused by the real essence of

this object. Because 'green' has not been caused by the real essence of water, which is H<sub>2</sub>O, we cannot say that my WATER representation in this case has tracked real water (even though it happens to causally covary with it).

I will not enter into a discussion about the plausibility of Prinz's thesis about tracking members of natural kinds. Rather, I am interested solely in the extent to which it can shed light on the relationship between emotions, values, and bodily changes. Prinz believes that there is a straightforward analogy between the two: just as we track natural kinds by registering their appearance properties, we track values by registering what they feel like. Let us compare the tracking of water and the tracking of offensiveness to see how exactly the analogy is supposed to run. The real essences in these cases are H<sub>2</sub>O and 'real offensiveness', the nominal essences are 'transparent, liquid, etc.' and 'feeling hot, having an increased heart rate, etc.', and whereas we track water—the stuff that contains H<sub>2</sub>O—by registering that there is something transparent and liquid, we track offensive remarks—those remarks that are really offensive—by registering that our cheeks are flushed and our heart rate has gone up. In the water case, we can say that a mental representation WATER refers to something that is indeed water if (NC) holds. Thus, in principle we should be able to say a mental representation OFFENSIVE refers to something that really is offensive if (NC) holds there too. And (NC) holds if and only if OFFENSIVE is reliably caused by objects that are really offensive, and if these objects cause OFFENSIVE in virtue of being really offensive.

I have set out how the analogy between tracking members of natural kinds and tracking values works according to Prinz. What might strike one as intuitively odd is the notion of a real essence in the case of values. What does it mean to say we track *real* offensiveness, or *real* danger, or *real* beauty? In the natural kind cases, we find it easy to believe that there are real essences, because the natural sciences tell us that they exist, and what they are. But in the value cases, no one has told us yet what exactly real offensiveness consists in. Intuitively, I think, many people would say that offensiveness is something we project onto certain remarks and behaviours, rather than something that is there in the world and that we can track. In other words, our feelings alone are all there is to an episode of anger, there is nothing over and beyond it that we track. Raffaella De Rosa raises a similar worry about Prinz's analogy in a

recent article, where she wonders what the real essence of humour might be:

In discussing the concept HUMOUR and the property it stands for, Prinz writes 'the property of being humorous does not look like anything but it is instantiated in various perceivable things: jokes ... garish make-up ... We can track properties by their instantiations' (p. 170). But, I wonder, in what sense in this case are appearance properties tracking something over and above (or, one might say, beneath and below) the appearance properties themselves (as in the case of GOLD)? There is no nominal/real [essence] distinction here to appeal to. But if this is the case, then there are at least some concepts that do not fit Prinz's theory (and the examples could be multiplied) (De Rosa 2005, 600).

There is then a worry that Prinz's analogy goes too far. Whereas in the case of water there is a real essence we track by registering appearance properties, there might not be much reason to think that there is a similar real essence of values that we track by registering bodily changes.

It is possible, however, that we just have to dig a little deeper to come to an understanding of the real essences of values. This is what I want to do in the next couple of sections: I will look at two possible candidates for the real essence of values, and see how plausible they are. In a recent article, David Pugmire examined Prinz's notion of real essences in the context of emotions and values as well, and developed and dismissed one candidate. In the next section, then, I will outline Pugmire's objection, and argue that Prinz is able to meet it. In the section thereafter, I will discuss what I believe Prinz himself would take to be the real essence of values, which I find very promising. Even though the obscurity around the real essence of values is then lifted, I will, however, develop an objection to his overall account that comes from quite a different angle.

## 2. The Pugmire objection

In his article 'Emotion and Emotion Science' David Pugmire analyses Prinz's distinction between real and nominal essences in the case of emotions, which he—justifiably—accuses Prinz of having left opaque. Pugmire tries to fill in the holes, and answers the question what it could possibly be that is tracked by our registration of bodily changes thus:

The real [essence] tracked by these [bodily] manifestations that we find in ourselves are the silent, internal neurochemical perturbations of the central nervous system. That is how the analogies of the relation of furriness to canine DNA or of sunshine to nuclear processes in the solar core translate for the relation of nominal to essential content of emotion events (Pugmire 2006, 15).

Pugmire believes that Prinz is committed to saying that the real essence of values are some “silent, internal neurochemical perturbations of the central nervous system”. Obviously, this strikes one as very counterintuitive, if not to say absurd. Surely, what makes a remark offensive is not some micro events in the brain, but—well, that it is really offensive. What then makes Pugmire think that Prinz is committed to saying that it is micro events in the brain?

Pugmire arrives at his conclusion in the following way: in order to identify the real essence of a value, he traces back the causal chain that results in our having a certain pattern of bodily changes. If we look again at the two figures, the plausibility of this strategy becomes clear. It is possible to track members of natural kinds by their nominal essences because these nominal essences are caused by their real essences. It is possible to track dogs by their nominal essences ‘furry, barking, and wagging its tail’, because these nominal essences are caused by their real essence, the dog genome. It is possible to track gold by its nominal essences ‘bright, yellow, and malleable’, because these nominal essences are caused by its real essence, AU. Finally, it is possible to track water by its nominal essences ‘liquid, transparent, and not tasting of much’, because these nominal essences are caused by its real essence, H<sub>2</sub>O. It seems fair enough then that Pugmire, in order to identify, for example, remarks that are really offensive, traces back the causal chain that results in the nominal essences of the offensive remark, that is, patterns of bodily changes. And the element he finds at the end of this causal chain is micro events in the brain. Has he then successfully reduced Prinz’s argument to absurdity?

I do not think so. Prinz is able to defend himself against Pugmire’s charges in the following way. In ‘The Duality of Content’, he discusses a potential problem with his account of tracking members of natural kinds, the solution to which I believe will help us here. Prinz calls this objection the horizontal problem. The horizontal problem arises if one takes (NC1), that is causal covariation, to be a sufficient condition for something to be the object to which a mental representation refers. After

all, it seems that there are many elements that causally covary with a representation of X. How do we know which one *is* X? Let me illustrate the horizontal problem with an example Prinz gives: in the case of the mental representation DOG there are many objects on the causal chain leading from the dog to the representation DOG: there is the creature with the dog genome, but there are also the light waves that hit my eye, the retinal image I form, and whatever else goes on in my brain after that. All these objects fulfill (NC1), that is, they all causally covary with my DOG representation. If (NC1) was sufficient for something to be the object to which a mental representation refers, then the retinal image I form before I come to represent a dog should qualify as that to which my DOG representation refers. But surely this is not the case. DOG representations refer to dogs, not to retinal images. Prinz ensures that we can solve the horizontal problem in this way by introducing (NC2). Not only does something have to causally covary with a particular mental representation in order to count as object to which it refers, but it has to causally covary in virtue of containing the particular real essence. And retinal images and DOG representations do not causally covary in virtue of the retinal image's containing the real essence of dogs. Instead, they causally covary because retinal images themselves are caused by objects containing the real essence of dogs. Hence, retinal images are not what mental representations refer to, even though they fulfill (NC1).

Pugmire's objection can now be answered in a similar way. Micro events in the brain might causally covary with OFFENSIVE representations, but they do not do so in virtue of containing, or being, the real essence of offensiveness. We can see this by considering the following case: say I misunderstood, in whatever way, a remark uttered by you. It will cause the same micro events in my brain that a truly offensive remark would have caused, and lead to the same set of bodily changes, and to my representation of your remark as offensive. Has my representation OFFENSIVE picked out something truly offensive though? No! And that is because your remark did not contain the real essence of offensiveness, which therefore must be something other than a characteristic set of micro events in the brain. Prinz is thus able to defend himself against Pugmire's objection: micro events in the brain only causally covary with OFFENSIVE representations, just like certain retinal images covary with DOG representations. But just as these retinal images do not

causally covary with DOG representations in virtue of containing the real essence of dogs, so do certain micro events in the brain not causally covary with OFFENSIVE representations in virtue of containing, or being, the real essence of offensiveness. They fail to fulfill (NC2).

While Prinz might thus be able to deny that he is committed to saying that certain micro events in the brain form the real essence of values, we still do not know what else forms them. We are back at the drawing board. What does it mean to be *really* offensive? I will turn to the second candidate for the real essence of values, the one I believe Prinz himself would put forward if asked. After that I will discuss a further objection to his overall account.

### 3. The objection from function

Prinz might be able to meet Pugmire's objection, but in order for his account to gain plausibility, his account still needs a defensible candidate for what the real essence of values might be. The candidate I believe he would put forward is this: based on the work of Richard Lazarus, Prinz defines values as core relational themes. Loss, danger, and offensiveness are all examples of core relational themes. They are relational in that they consist of a relation between something I care about, and how the world is. Loss, for example, can be the relation between my beloved grandmother, and the world that does not accommodate her anymore. Danger is the relation between my wellbeing, and a situation in the world that threatens my wellbeing. Offensiveness is the relation between my respectability, say, and an action or a remark that shows a lack of respect. In each of these cases we have a relation between something I care about and the way the world is. Whatever it may be that I care about, if it is not accommodated by the world anymore, I will be sad and find the cause of its removal a loss; if it is threatened, I will be afraid and find the cause of this threat dangerous; and if it is disrespected, I will be angry and find the cause of this disrespect offensive. It is relations of this kind that emotions are meant to track. Thus, it is relations of this kind that form the real essence of values.

Now that we seemingly have the last piece of the puzzle, let us try to put it together using the example of offensiveness. Prinz's full account of how we come to represent something as offensive has to run roughly like this:

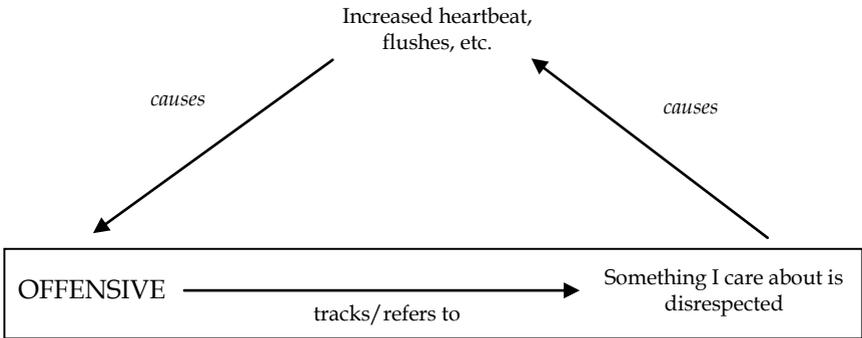


Figure 3: Tracking offensiveness by registering bodily changes.

In an episode of anger, I come to represent a remark or an action as offensive by registering that my heartbeat has increased, my cheeks are flushed, etc. My mental representation OFFENSIVE refers to something that really is offensive if it causally covaries with remarks or actions that disrespect what I care about (NC1), and if these remarks or actions cause OFFENSIVE in virtue of being really offensive, i.e. in virtue of actually disrespecting something that I care about (NC2).

I believe that relations between objects we care about and the world are very good candidates for what values really are. The definition of values as core relational themes has an intuitive appeal to it: anger seems to be indeed directed at a relation between something I care about and the remark that slights it. As soon as this relation is removed, my anger will disappear. If I do not care about something, then a remark that slights it will leave me cold. Similarly, if there is something I care about that indeed has been slighted, then my behaviour will be considered somewhat silly if I continue holding a grudge once the slight has been taken back. Values come as these relations: something is only valuable if it is cared about, and its specific value will be determined by the kind of situation it is in. Thus, I accept what Prinz’s would most likely suggest as that which makes a value a value, that it, its real essence.

But one might, at this point, raise an objection in Pugmire’s spirit. Core relational themes might be a very plausible candidate for what values essentially are, much more so than micro events in the brain; however, it is central to Prinz’s picture that there is a causal link between real essence and nominal essence. While it was easy to understand the

causal link between micro events in the brain and bodily changes, it seems much less easy to understand the causal link between core relational themes and bodily changes. After all, there was a reason why Pugmire suggested micro events in the brain as a candidate for the real essence of values: they are at the end of the observable causal chain leading up to bodily changes. Taking core relational themes as the ultimate cause seems—at least in this respect—far more speculative and vague. I believe that Prinz would meet these worries by pointing to the bigger teleosemantic picture, which serves as the background to his theories. It might not be entirely clear *how* core relational themes come to cause the bodily changes that represent them, but *that* they do is supported by the fact that it is the function of bodily changes to track them. We undergo the bodily changes we do because it is their function to inform us that something is the matter: we're in danger, offended, attracted. And if it is their function to inform us thus, bodily changes must be caused by what they inform us about. For how could they inform us that we are in danger, if they were not reliably caused by instances of danger?

It is precisely this thesis that it is the function of bodily changes to inform us of value relations that I want to take issue with. At least, I believe, that informing us of value relations is not the essential function of bodily changes. They have a second, and to my mind much more essential, function, namely to prepare and motivate us for action. So while Prinz dwells on the *descriptive*, the tracking, side of the bodily changes involved, I think he neglects their *directive* side. Consider the following three examples, which to my mind show that action-directing is an essential function of the bodily changes involved in emotions. When we come to represent something as offensive, we feel motivated to retaliate. The flushes and the heat we are experiencing when angry prepare us to do something about the offense. Were we all calm and cool when somebody insulted us, we would not be motivated and prepared to do something to set things right. However, we could still form the judgement that somebody has offended us. Furthermore, when we come to represent something as dangerous, we feel motivated to run away or attack. The increase of speed of our heartbeat might be said to enable our bodies to cope with running away quickly. Again, were we all calm and collected, we could still form the judgement that we were in a dangerous situation, but we would not be prepared to react properly. Finally, and more controversially, when we come to represent something as beautiful, we

are disposed to protect it. We cannot represent something as beautiful and remain indifferent to its welfare, be the object a painting, a landscape, or a person. The hormone rushes we undergo might be said to provide us with a deep motivation to look after this object that causes them. Once more, were we all calm and collected, we might still be able to appraise the beauty of an object, but would be in no way disposed to show this in our behaviour.

Prinz would disagree in that he would deny that the bodily changes involved in emotions are inherently motivating. According to him, emotions and motivations are two different mental states: "Motivations are action commands that are pushed or pulled by affective states. Motivations are often pushed or pulled by emotions. But when emotions cause motivations, those motivations never count as constitutive parts of emotions. The two constructs are ... closely entwined, but independent" (Prinz 2004, 196). Thus, emotions can cause motivations, but motivations are not something that an emotion ever consists in. An emotion, according to Prinz, is an embodied appraisal that can—or cannot—lead to a preparation and motivation to act. Based on my previous examples, I want to contradict this. Since bodily changes are essential ingredients of emotions, and their main function is to motivate us, I believe that motivations are essential ingredients of emotions. We do not undergo two separate sets of bodily changes, one which represents a value relation to us, and another which prepares us to react. There is only one set, and its essential function is action preparation. As an argument for this claim I want to appeal to the phenomenology of an emotional episode: it does not feel like we undergo two separate affective states when, for example, we are angry and motivated to retaliate. My increased heart rate, my flushes, etc. are all part and parcel of my motivation to do something about the offense. They do not just tell me that I have been offended: they are one step ahead; they are preparing me to react. Also, when I am afraid, I do not first register that there is something dangerous via registering my trembling hands and quickened heartbeat, and then I undergo a second, separate, affective state that motivates me to run away. Instead, it seems that there is one pattern of bodily changes only, and it essentially prepares and motivates me to run away. I will call the fact that Prinz denies the bodily changes we undergo in an emotion their motivational dimension Prinz's 'motivation problem'.

Assuming I am right in claiming that the function of the bodily change we undergo in an emotional episode is to prepare and motivate us for action, and that Prinz does have a problem because he does not account for this, does it follow that Prinz's claim that bodily changes track values is wrong? It does not necessarily mean this. For it is possible that bodily changes *both* track values *and* motivate action. Their function could be to describe and direct at one and the same time. This view is in fact held by at least two philosophers, Andrea Scarantino (2005) and Carolyn Price (2006). They interpret emotions as pushmi-pullyu representations—a term coined by Ruth Millikan—that is, as states that describe and direct. In the next section, I will outline how we can understand emotions as pushmi-pullyu representations according to Scarantino and Price, and analyse whether this could be a solution for Prinz's motivation problem.

#### 4. Emotions as pushmi-pullyu representations?

A pushmi-pullyu representation is a representation that is descriptive and directive at the same time. One of Millikan's original examples is the dance bees perform to indicate a new source of nectar. This dance does not only inform the fellow bees that there is a new source of nectar nearby, it also aims at guiding their behaviour, for they are meant to fly there directly. One single act fulfils two functions: it describes a location, and motivates others to fly there. Another example of Millikan's is the utterance "women don't do this kind of thing". This one utterance is both meant as a statement about how the world is, and an action direction.

Millikan herself does not apply the concept of pushmi-pullyu representations to emotions, but Scarantino and Price have done so. Scarantino claims about emotions that "their purpose is to *guide behavior according to how the world varies*" (Scarantino 2005, 261), while Price states that "an emotional appraisal is supposed to be produced in a particular type of situation, and it is supposed to ensure that the subject's response is appropriate to a situation of that type" (Price 2006, 216). Thus, an emotional representation is meant to track particular situations in the world, and guide our behaviour appropriately. It is meant to describe value relations, and direct our behaviour accordingly.

Could this be a solution to Prinz's motivation problem? It could, if one identified patterns of bodily changes as the central element that does

the describing and directing. Interestingly, however, Scarantino would want to deny this. He believes that appraisals are the central elements in emotions, and his definition of an appraisal—unlike Prinz’s—does *not* include registrations of bodily changes. Scarantino believes that appraisals range from modular appraisals, in which there is a dedicated neural pathway in which the information transmitted is encapsulated, to central appraisals, in which there is no dedicated neural pathway, that is, in which the information transmitted is penetrable. Modular appraisals are found in humans and animals alike, whereas central appraisals are unique to humans (Scarantino 2005, 225). I do not think it is necessary to go into any further detail about Scarantino’s theory of appraisal; it suffices to say that they are quite independent from registrations of bodily changes. Put differently, this means that we can come to represent value relations without the help of registered bodily changes. On Scarantino’s picture, bodily changes enter the picture only much later in the process, when the body prepares for action. The appraisal of something as offensive, dangerous, or beautiful, has nothing to do with them.

Whether Scarantino is right or not, one thing should be noted: an interpretation of emotions as pushmi-pullyu representations is not a straightforward affair. Pushmi-pullyu representations are characterised by one central state that both describes and directs. According to Scarantino, this central state seems to be an appraisal made without the help of any registrations of bodily changes. It is an appraisal that represents value relations to us, and motivates us to act. But if we want to solve Prinz’s motivation problem by turning his theory into a pushmi-pullyu account, this central state will have to be a registration of bodily changes. Prinz is certain that bodily changes describe, that is, represent value relations. To overcome the motivation problem we would have to ascribe the function of directing to bodily changes also. But is this correct? Which one is the central state that does the describing and directing, Scarantino’s appraisals or Prinz’s registrations of bodily changes?

I do not believe that there is a simple answer to this question. Rather, I believe that both Scarantino and Prinz get some things right, and some things wrong. The truth lies somewhere in the middle. Scarantino is right, I believe, in identifying the appraisal of something as offensive, dangerous, etc., as a process that can be quite independent of any bodily changes. We assess that something we care about stands in a certain relation to the world by means of our ordinary cognitive system—

simply do not need bodily changes to do this job. By this I do not mean to suggest that we are never informed about a value relation by patterns of bodily changes. For example, the bodily changes involved in fear are often experienced first, before a conscious appraisal of something as dangerous has been made. These bodily changes are the first element we become aware of, and it seems as though they inform us that something dangerous is around. But this is only the case because we become *aware* of them first – there will have been a preceding appraisal that took place on a subpersonal level, something Scarantino calls a modular appraisal. Thus, the fact that we become aware of the dangerous situation only by registering the bodily changes we are undergoing is a side effect of these bodily changes; it is not what they are meant to do. We have already been informed of the danger in a different way; a way to which we have no conscious access. The bodily changes are a result of this information, not a means to transmit it. They are meant to prepare us to run, or attack, or whatever is appropriate in that situation.

Then again, I believe that Prinz is right in stressing the centrality of bodily changes, even if for the wrong reason. The experience of patterns of bodily changes is central to an emotional experience. A mere appraisal is insufficient to make us afraid. Being afraid essentially consists in *feeling* afraid. Only when our heart races and the anxiety begins will we experience true fear. But we do not need these patterns of bodily changes to inform us of value relations, rather we need them to quickly prepare us for action. A calm and collected appraisal of something as dangerous is insufficient to prepare us for an immediate and effective fight or flight reaction. We need the bodily changes involved to make us run, attack, defend, or protect. They account for the motivational, and the directive, side of emotions. Furthermore, they seem to be – at least to a certain extent – independent from the appraisals that often precede them. Undoubtedly, bodily changes are often caused by appraisals, but this is not necessarily the case. Drug-induced bodily changes motivate just as much, if not more, than bodily changes caused by appraisals. Consider a person who has taken a drug that prompts all the bodily changes characteristic of fear. He will feel afraid, and be motivated and prepared to run away, even though he has not appraised anything as threatening something he cares about.

All this suggests that appraisals of value relations and motivating patterns of bodily changes are two separate states. Emotions seem to

have two functions: to inform us of value relations, and to motivate us to take action. I believe that each of the two states fulfils one function: appraisals inform us of value relations, and bodily changes motivate us to act. Because the two functions are fulfilled by two different states, I think it is inappropriate to speak of emotions as pushmi-pullyu representations. After all, pushmi-pullyu representations are characterised by one single state accomplishing both directive and descriptive functions. Instead, I propose the following model of how value relations, appraisals thereof, and bodily changes are connected. Value relations are relations between objects I care about and various situations in the world. These relations can hold without my being aware of them. But if I am aware of them, I will make an evaluative appraisal. If my dog, who I care about, is threatened by the much bigger neighbour's dog, I will appraise this as a dangerous situation. This appraisal will prompt various bodily changes in me, which motivate and prepare me to take action. I will be very alert, and ready to step into the situation, trying to keep the bigger dog away from my dog. The emotion of fear I am experiencing comprises *both* the appraisal *and* the motivating pattern of bodily changes. An emotion proper is one in which an evaluative appraisal is followed by an appropriate set of bodily changes. This, probably most often, happens so quickly that we will not differentiate between the two different states. But there are two different states at play, and neither on its own is sufficient to constitute an emotion. Appraisals alone are not emotions. A calm and collected appraisal of a remark as offensive does not amount to anger. Similarly, registrations of bodily changes alone, though perhaps experienced as emotions, are not emotions proper either. The drug-induced person is not *properly* afraid. This is because his feelings have not been caused by an evaluative appraisal of an object as threatening something he cares about. An emotion proper will begin with a conscious or unconscious appraisal of a value relation, and followed by a set of bodily changes that prepares us to take the appropriate action.

Obviously, there is a lot more that can and should be said about this model. I will end, however, by putting forward a few concluding remarks on its advantages, compared to Prinz's account in particular, and more generally. First, I will briefly summarise the main points made in this paper. In section one I gave a detailed account of how Prinz believes that emotions represent values by registering bodily changes. I concluded that the notion of a real essence of values remained obscure. In the

second section I discussed Pugmire's suggestion for what the real essence of values might be, his objections to it, and how I believed Prinz could meet these objections. In section three I outlined what I took to be Prinz's idea of the real essence of values, that is, core relational themes, and argued that this was a plausible account of the essence of values. Less plausible, however, was the background assumption on which Prinz's theory stands: that it is the function of bodily changes to represent such core relational themes. I argued that the first and foremost function of bodily changes in emotions was to motivate and prepare us for action. Prinz does not account for this, a fact which I called Prinz's 'motivation problem'. In section four, I considered a possible solution to Prinz's motivation problem: the construction of emotions as pushmi-pullyu representations. I concluded that this was not promising, as emotions seemed to consist of two central states rather than one: appraisals and patterns of bodily changes. Thus, I constructed my own model based on these findings: emotions are compounds consisting of an appraisal of value relations, which causes a pattern of bodily changes that is meant to motivate and prepare us for action.

The major attractions of my model are these: for once, it bypasses some of the problems that Prinz runs into. It accounts for the fact that the bodily changes we undergo in emotions are meant to prepare us for action, which solves the motivation problem. Furthermore, it is more intuitive in that it respects the phenomenology of emotions: we appraise value relations, which cause bodily changes, which motivate us to take appropriate measures. It strikes me that this is what introspection reveals to us as going on when we experience an emotion, rather than that we undergo bodily changes first, which in turn represent value relations. Finally, it keeps the major advantage of Prinz's theory, that is, it accommodates both the bodily and the evaluative side of emotions: as I explained in the introduction, this is a major criterion for any theory of the emotions.<sup>2</sup>

<sup>2</sup> I would like to thank Tom Baldwin, Christian Piller, Dorothea Debus, Robin Dennis, Rachael Wiseman, Bob Clark, the participants of the conference "Emotions and Intentionality", held in Prague in April 2008, as well as the research students of the philosophy department at the University of York, for many valuable comments.

Department of Philosophy  
University of York  
ed503@york.ac.uk

## References

- DEONNA, J.A. (2006): Emotion, Perception and Perspective. *Dialectica* 60, 29 – 46.
- DE ROSA, R. (2005): Prinz's Problematic Proxytypes. *The Philosophical Quarterly* 55, 594 – 606.
- DE SOUSA, R. (1987): *The Rationality of Emotions*. Cambridge MA: MIT Press.
- DÖRING, S. (2007): Seeing What to Do: Affective Perception and Rational Motivation. *Dialectica* 61, 363 – 394.
- JOHNSTON, M. (2001): The Authority of Affect. *Philosophy and Phenomenological Research* 61, 181 – 214.
- MCDOWELL, J. (1998): Values and Secondary Qualities. In his *Mind, Value and Reality*. Cambridge MA: Harvard University Press.
- MILLIKAN, R.G. (1995): Pushmi-Pullyu Representations. In TOMBERLIN, J. (ed.): *Philosophical Perspectives*, Vol. 9. Atascadero, CA: Ridgeview Publishing Company, 185 – 200.
- PRICE, C. (2006): Fearing Fluffy: The Content of an Emotional Appraisal. In MACDONALD, G. – PAPINEAU, D. (eds.): *Teleosemantics*. New York: Oxford University Press, 208 – 228.
- PRINZ, J. (2004): *Gut Reactions: A Perceptual Theory of Emotion*. New York: Oxford University Press.
- PRINZ, J. (2000) The Duality of Content. *Philosophical Studies* 100, 1 – 34.
- PUGMIRE, D. (2006): Emotion and Emotion Science. *European Journal of Analytic Philosophy* 2, 7 – 27.
- SCARANTINO, A. (2005): *Explicating Emotions*. Ph.D. thesis. Pittsburgh: University of Pittsburgh.